

Evaluating Evaluation

Increasing the Impact of Summative Evaluation in Museums and Galleries

**Maurice Davies and Christian Heath
King's College London**

November 2013

maurice.davies@ntlworld.com

christian.heath@kcl.ac.uk

Contents

Executive summary and key recommendations	3
Introduction	8
The problem	
Types of evaluation	
Scope and method	
Potential benefits of summative evaluation	
1 Methods: data collection and analysis	12
2 Concepts and models of the visitor	16
Summary of sections 1 and 2	19
3 The organisational and institutional framework	21
Summary of section 3	28
4 Conflicting purposes	30
5 Increasing the impact of summative evaluation and key recommendations	31
6 Overall conclusion	36
Acknowledgements	37
Appendix 1: Suggestions for improving the impact of summative Evaluation	38
Appendix 2: An overview of some Principal Findings of Summative Evaluation	44
Bibliography	67

Executive Summary

The starting point of the Evaluating Evaluation project was our impression that despite the substantial resources that are spent on the summative evaluation of museums and galleries the research has little impact and largely remains ineffectual.

With support from the Wellcome Trust and the Heritage Lottery Fund, we set out to see why this seems to be the case and to explore whether there are ways in which the impact of summative evaluations on knowledge and practice might be enhanced. To this end, we reviewed a substantial number of reports, undertook a range of interviews and held two colloquia that brought together a range of experts, including those who undertake evaluation, those who might have an interest in the findings and those who commission the research. As well as identifying a number of issues that undermine the impact of evaluation and making a series of recommendations, we also summarise the more general findings and implications that emerge from this wide-ranging body of applied research. These findings are set out in appendix 2 and summarised there.

We chose to focus on summative evaluations. In most cases they are the most substantial studies that examine how visitors respond to museums and galleries and are a requirement of most externally funded projects that benefit from public, lottery and charitable money. They also result in published or potentially publishable reports. We restricted our attention largely to summative evaluations of permanent displays in museums and galleries, but we included a small number of summative evaluations of temporary exhibitions, historic buildings and science centres (especially where the evaluations were distinctive, perhaps because they included new methods or less common findings, or were of higher quality than usual). We did not look in any detail at summative evaluations of activities, such as education programmes or community projects.

Overall conclusion

As set out in section 6, after a year of reading and discussing summative evaluation with a broad range of practitioners, our feeling is best expressed as disappointment. Disappointment that all the energy and effort that has been put into summative evaluation appears to have had so little overall impact, and disappointment that so many evaluations say little that is useful, as opposed to merely interesting. (Having said that, we found some evaluation studies that do begin to get to the heart of understanding the visitor experience.) With some notable exceptions, summative evaluation is not often taken seriously enough or well enough understood by museums, policy makers and funders. The visibility of summative evaluation is low. Too often, it is not used as an opportunity for reflection and learning but is seen as a necessary chore, part of accountability but marginal to the work of museums.

The impact

Our investigation found that it is rare for a museum to act directly on the findings of summative evaluation by making changes to the gallery that was evaluated. It is

more common, but still unusual, for a museum to fully apply the findings of evaluation in the development of subsequent galleries. This can be down to an absence of internal systems to use the knowledge generated by evaluations, but – practically speaking - in cases where exhibitions or galleries are very varied, rather than following a fairly consistent design, the findings of previous evaluations may not be very relevant to subsequent work. There are, of course, exceptions. We found a few examples of museums systematically using evidence and recommendations from a series of evaluations to more generally inform the development of subsequent gallery and exhibition spaces. In these museums evaluation has a significant impact on practice.

We also found that despite the substantial number of summative evaluation studies that have been undertaken over the last decade or so, they have made a limited contribution to practice more generally and our overall understanding of visitor behaviour and experience (for more detail see appendix 2).

The relative lack of impact of summative evaluation derives in part from methodological and conceptual issues, but we found that most significant is the organisational and institutional framework in which summative evaluation is undertaken. As set out in sections 1-3, the reasons include:

1 Variable **methods** of data collection and analysis are used in museum evaluation studies, which do not easily enable comparative analysis or the establishment of a more general corpus of findings and recommendations:

- evaluations are primarily structured to address the specific aims and objectives of a particular project
- individual evaluations use a diverse range and combination of methods to collect and analyse data
- data is analysed and presented in varied ways
- there can be methodological shortcomings and analytic limitations in the application of particular techniques

2 Evaluations adopt a range of distinctive **concepts and models** of the visitor, including for example theories of visitor motivation:

- different models and conceptions of visitors and their behaviour underpin different evaluation studies (or the work of different organisations) and thereby undermine comparative analysis
- most summative evaluations prioritise the individual and tend to neglect the importance of social interaction in how visitors behave in and experience museums and galleries
- what visitors actually do at the ‘exhibit face’, the actions and activities that arise at exhibits, is given relatively little analytic attention in summative evaluation

3 The **organisational and institutional environment** in which evaluation is commissioned, undertaken and disseminated can undermine the impact of the findings and the development of a body of shared knowledge and expertise. The

organisational and institutional framework undermines the possibility of preserving and transferring knowledge of good and poor practice across projects within institutions and sharing the findings of evaluations between organisations. The constraints placed on summative evaluation include:

- the ways in which studies are tailored with regard to the specific requirements of particular initiatives
- the necessity to provide a largely positive assessment of the project's outcomes and the achievement of its goals
- the impracticality and expense of undertaking substantial 'remedial' changes to an exhibition, gallery or even exhibit
- the lack of opportunities or mechanisms to share experience and establish a repository of knowledge and practice both within and across institutions
- the limits placed on access to reports outside, and - in some cases – within, particular organisations
- the temporary, project-specific, organisation of design and development teams especially in the case of major initiatives
- the use of external design consultants, who temporarily bring their expertise and knowledge to bear on the development of exhibitions and exhibits but who are not normally included in the evaluation process
- the detached or marginal institutional position of some of the individuals or organisations who undertake evaluation
- the varied purposes of, and expectations of, summative evaluation

This latter point is crucial: summative evaluations can be subject to different, often **conflicting purposes**, as discussed in section 4. These include:

- reflection on the particular project and learning within the project team
- learning and knowledge more widely - within the organisation and across the sector
- monitoring and accountability to both the institution and the funder
- advocacy, for example in support of bids for future funding

Improving the impact of evaluation

In section 5, we summarise ideas for improving the impact of evaluation, based on observations of successful organisations and suggestions made at two colloquia and in discussions with individuals (fuller details are given in appendix 1); we also make a number of recommendations to museums, to evaluators and to funders. We hope all three groups, perhaps working together, will find ways to act on these recommendations. Perhaps a starting point could be a concerted push by the evaluation community (that is evaluators themselves and funders and museum staff who recognise the benefits of evaluation) to demonstrate the value of evaluation. We are pleased that the Visitor Studies Group is intending to play a role in taking forward the findings of the Evaluating Evaluation investigation, for example by organising seminars and discussions.

Key Recommendations

A) Recommendations for Museums

A1 The first task of a museum requiring summative evaluation is to specify what it is intended to achieve and how it will be used. Evaluation briefs should be clear about the use(s) to which the evaluation will be put, how it will relate to other evaluation and audience research, and how it will be disseminated within the museum and possibly more widely.

A2 Embed evaluation in the museum. Have an evaluation framework and include evaluation as a key part of other plans and strategies. Allocate responsibilities for evaluation and audience research. Have mechanisms in place for preserving and disseminating knowledge from evaluations among staff and across successive projects.

A3 Disseminate the findings and implications of summative evaluations across the broader community including museum staff, evaluation consultants and designers, by publishing reports, or digests of key findings, and organising or contributing to workshops, seminars and conferences.

A4 Identify overarching research questions to inform individual pieces of summative evaluation. Consider collaborating with other museums to establish a common set of research questions and secure comparable data.

A5 Enable and support comparative analysis and the synthesis of findings from a range of summative evaluation studies through adopting at least some common themes, adopting similar research methods and building an accessible data corpus and comparable analytic framework. Consider raising funding, or building partnerships, to enable this to happen.

A6 Plan for change. Build in time and money for remedial change, to act on the findings of an evaluation.

b) Recommendations for Evaluators

Some of these recommendations apply to individual evaluators; others to evaluators working together, possibly via an organisation such as the Visitor Studies Group

B1 Raise awareness of the benefits of evaluation across the sector. Explore ways of increasing museum practitioners' skills in commissioning and understanding evaluation. Training or good-practice advice on briefing could be particularly

effective. Showcase examples of best practice, including examples where evaluation has had a significant impact on practice.

B2 Share the findings of evaluations. When possible, publish reports, or at least digests of key findings. Contribute to seminars and conferences.

B3 Enable and support the comparison and synthesis of findings from a range of summative evaluation studies. Consider raising funding, or building partnerships, to enable this to happen.

B4 Devise overarching research questions and encourage museums to support summative evaluations that contribute to the research questions.

B5 Consider standardising some aspects of data collection and methodology to aid comparability of evaluation studies.

B6 Exploit contemporary methodological developments in the social sciences and consider using more innovative ways of collecting and analysing data.

c) Recommendations for Funders

C1 Be precise in specifying what summative evaluation is intended to achieve.

C2 Engage with grant-recipients' evaluation work: set high expectations of transparency and honesty; ask questions of the report.

C3 Encourage (or require) museums to act on the findings of summative evaluation and provide funding to allow changes to be made to galleries. Consider requiring plans for acting on the findings of each evaluation.

C4 Raise awareness of the potential benefits of evaluation across the sector. Explore ways of increasing museum practitioners' skills in commissioning and understanding evaluation. Training or good-practice advice on briefing could be particularly effective.

C5 Encourage dissemination and sharing of the findings of summative evaluation studies, and synthesising the findings of a range of evaluation studies. Consider taking responsibility for ensuring these activities happen.

C6 As a quality control and audit, directly commission (and publish) fully independent evaluation of a sample of projects.

Evaluating Evaluation: Increasing the impact of Summative Evaluation

Introduction

The problem

Over the past decade or more, a great many new galleries in museums have been subject to summative evaluation, resulting in a substantial amount of observations and findings and yet their influence on practice seems to be limited. It is rare for a museum to act on the findings of summative evaluation by making changes to the gallery that was evaluated; when instances are found, they can sometimes be almost idiosyncratic, rather than the consequence of systematically applying the key findings of the evaluation. There are some examples of the findings of summative evaluation influencing the design of subsequent galleries and museum spaces in the same institution, but this is rarely done systematically. A typical change is amending directional signage. One evaluator said that of all types of evaluation, summative evaluation is the hardest to make use of and the least likely to have an impact, unless it is within an evaluation framework (King, 2012b).

More surprising perhaps, is that the insights and observations generated by these numerous and in some cases quite extensive summative evaluations, have not formed a robust body of concepts and findings that more generally inform museum practice and theory - in particular in the area of exhibition design and development. This view is shared by a review of informal science learning, 'there was little evidence from the literature review or the stakeholder interviews to suggest that learning from evaluation work is commonly shared across the sector to inform activities more widely.' (GHK Consulting, 2012, p. 38)

With the help of funding from the Wellcome Trust and HLF, we set out to see why this seems to be the case and whether there might be ways that summative evaluations could have a greater impact on practice.

We also reviewed a substantial number of (mainly summative) evaluations to collate information about what evaluations can tell us generally about the behaviour and experience of visitors. Some of our rather limited generalizable findings can be found in appendix 2.

Types of evaluation

The UK's main funder of museum developments says evaluation is 'reflection on the project... looking back on what you did, and finding out how well it worked... We think it leads to better projects [and] it will stand the heritage sector in better stead if it is able to draw on a richer bank of

evidence to demonstrate why heritage matters and what it can achieve. Evaluation has two purposes – one is about proving, the other is about improving. Proving means demonstrating that change is actually taking place... When viewed as an *im*-proving exercise, then evaluation needs to be part of a continuous process of learning and growth.’ (Heritage Lottery Fund, 2008, pp. 5-6)

With its standpoint of ‘looking back’, this is essentially a description of the role of summative evaluation. Conventionally, evaluation is often divided into three types: front-end, formative and summative:

‘Front-end evaluation occurs during the planning stages of a project or activity... before resources are committed to developing the project. It is used particularly during the planning stages of museum and gallery redevelopment... *Formative* evaluation occurs during the process or development of your project or activity. It may involve testing out a prototype or mock-up... [or] meetings to monitor and assess how well a project is progressing with a view to making amendments to the plan if needed. At this point you are considering the question, “are we getting to where we want to be?” *Summative* evaluation occurs at the end of your project or activity, when you may wish to assess the impact of the finished product... At this point, you are looking back at the project to ask questions such as “Did we meet our aims and objectives?” or “What worked well and what didn’t?”’ (East of England Museum Hub, 2008, pp. 8-9)

Some argue against a hard division between front-end, formative and summative evaluation and instead suggest evaluation is better as a continuous process, running alongside the project – a ‘continuum’ or ‘end to end process’ (Wilkins, 2012). This approach, in which an evaluator encourages reflection and improvement throughout a project, is relatively unusual, but can be found in projects funded by organisations such as the Paul Hamlyn Foundation or Esmée Fairbairn Foundation.

Scope and Method

We chose to focus on summative evaluations. In most cases they are the most substantial studies that examine how visitors respond to museums and galleries and are a requirement of most externally funded projects that benefit from public, lottery or charitable money. They also result in published or potentially publishable reports. Summative evaluation is also of interest because, notwithstanding the substantial resources that it requires, there is a lack of clarity about its purpose and evaluators themselves express dissatisfaction with it, telling us for example that summative evaluation is often the least interesting type of evaluation, especially when isolated from front-end and formative (Boyd, 2012b).

We restricted our attention largely to summative evaluations of permanent displays in museums and galleries, but we included a small number of summative evaluations of temporary exhibitions, historic buildings and science centres (especially where the evaluations were distinctive, perhaps because they included new methods or less common findings, or were of higher quality than usual). We did not look in any detail

at summative evaluations of activities, such as education programmes or community projects.

The material that forms the basis of this report is derived from various sources. First and foremost, we reviewed a substantial corpus of summative evaluation reports of various developments, both major and minor in a broad range of museums and galleries throughout the UK. Secondly, we organised two colloquia that brought together a range of practitioners, all of whom have an interest in evaluation, its findings and implications. These include representatives from funding agencies that expect evaluation, those who undertake evaluation as consultants as well those based in particular museums and galleries, curators, museum managers, educationalists, and designers who are involved in the development of major projects. We also included a small number of academic researchers who have a particular interest in evaluation. These colloquia provided a range of insights into the organisation, practicalities and conflicting demands of evaluation, and also helped provide access to a number of follow up interviews and discussions with particular individuals. We developed a draft report based on the analysis of the materials that derived from the review, the colloquia and the interviews, that was presented at the second colloquia to elicit comment and feedback. It was then finalised in the light of this additional information.

The potential benefits of summative evaluation

When done well and taken seriously, summative evaluation can bring many benefits. It has the potential to, for example:

- Increase understanding of visitor behaviour and engagement
- Provide information about the impact on individuals of engaging with a display
- Identify improvements that could be made to existing displays
- Provide findings that can inform the development of galleries to enhance the engagement of visitors
- Suggest areas for experimentation to see whether particular approaches or techniques improve visitor engagement
- Confirm, or legitimise, the views and experience of audience-orientated staff
- Reveal unexpected outcomes or impacts
- Stimulate reflection and learning by staff involved in a project
- Contribute to wider learning and knowledge about visitor behaviour and interaction and techniques to enhance impact
- Improve museum accountability and transparency

Why so little impact?

However, most summative evaluation fails to achieve most of these potential benefits and has a rather smaller impact, as one anonymous contributor to the 2nd Evaluating Evaluation colloquium wrote,

‘In terms of exhibitions, summative evaluation often reveals small areas where improvement can be made for the next exhibition. Small, iterative

improvements can make a significant difference – small margins are often important to the overall success of a display element. If we are aspiring to create the best displays we need to pay attention to detail – small margins matter. We might learn more from front-end and formative than summative – but summative is still important.’

Why does summative evaluation usually have only a marginal effect? There are a number of possible explanations for the apparent lack of impact of summative evaluation studies on museum practice and our understanding of visitor behaviour. These include:

- (1) the different methods, and variations in their application, used in museum evaluation studies that do not easily afford comparative analysis (and it is argued in some cases fail to meet the rigours required of a social science);
- (2) the use of distinctive theories of visitor behaviour and motivation by different studies (and consultant organisations) that provide a very specific focus for the research; and
- (3) the museum’s organisational environment that undermines the possibility of preserving and transferring knowledge of good and poor practice across projects within particular institutions and of sharing knowledge more widely.

Section 1 Methods: data collection and analysis

A broad range of quantitative and qualitative methods underpin data collection and analysis within summative evaluation. These methods reflect methodological developments within the social sciences and the changing nature of the disciplines involved in museum studies. For example, early studies of visitor behaviour largely arose within behavioural psychology and focused in particular on navigation patterns of individual visitors and the length of time spent at particular exhibits. Concepts such as 'dwell time', 'attraction' and 'holding power' permeate these studies and continue in many studies today. With the cognitive turn that arose in the social sciences over the last couple of decades or so, we have witnessed a growing emphasis on interpretation, experience and learning and a corresponding shift in methodological and conceptual commitments found within summative evaluation. These developments resonate with a socio-political agenda that has increasingly prioritised access, engagement and learning, an agenda that has had an important influence on the funding of new museum and gallery spaces.

Almost all summative evaluations use a mixture of quantitative and qualitative methods. The methods they use and the particular combination vary from project to project, but ordinarily they involve at least two or three of the following: surveys, face to face interviews (both structured and unstructured), focus groups, accompanied visits, gallery observation (both participant and non-participant), anatomies of museum visits, vox pops, Personal Meaning Maps (PMM), and audio and video recording. For example, the evaluation of All About Us at @Bristol (Gammon & al, 2011) gathered data using a combination of exit interviews, accompanied visits with families and school groups, and focus groups with parents and staff. Even a smaller scale project such as the evaluation of the Jameel Gallery at the V&A (Fusion, 2008) uses a combination of methods, in that case, tracking of visitors, face to face interviews, and PMM. In using a diversity of methods, evaluations routinely generate a range of data that allows investigation of a broad range of issues that may include for instance: the profile of visitors, the route and duration of their visit, their attitudes towards and attention paid to particular exhibits, information resources and the like, and whether there is evidence of learning.

From the point of view of an individual evaluation, it makes sense to gather a range of data that enables consideration of a wide diversity of issues. However, as different evaluations often use different combinations of methods it can prove difficult to compare and contrast the results between evaluations, sometimes even between those undertaken within the same institution. To make matters more difficult still, the same methods may be applied in very different ways, using different criteria with a distinctive analytic focus, and in consequence the findings generated by seemingly similar studies prove incomparable. In other words, the specific requirements placed on particular projects, by for example the institution or funding body, coupled with the methodological commitments of a specific evaluation team,

can serve to encourage a distinctive evaluation study that generates observations and findings that are highly specific to the project in question.

These difficulties are further exacerbated by virtue of the varied ways in which data is subject to analysis. Take, for example, data gathered through observation, focus groups and in-depth interviews. There are two ways in which this is typically subject to analysis: (i) events/opinions are categorised and a simple frequency count is used to assess and compare certain characteristics; (ii) quotes from visitors and/or field notes are used to illustrate and develop particular issues. What behaviours and events form the focus of the analysis and how they are categorised can vary significantly between different projects. Moreover, while selected extracts from field notes, interviews or focus groups can prove highly illuminating within the framework of a particular project, it can prove difficult to systematically compare the information to material taken from other studies.

While it may be difficult to compare and contrast results across distinct forms of summative evaluation, we found occasional examples where an institution is building a more general corpus of findings and practical strategies by utilising similar models and methods through a series of evaluations, using evidence and recommendations from them to more generally inform the development of gallery and exhibition spaces (see, for example, case study on p26).

It has also been suggested that in some cases evaluation studies fail to follow the strictures required in rigorous, social science research and thereby undermine the quality and usefulness of their observations and findings. Jensen suggests, for example, that the samples used in evaluation studies' surveys or tracking studies, can be unreliable, based on convenience rather than rigorous criteria, and that questionnaires are often poorly designed and bias particular responses (Jensen, 2012). With regard to observational research, we can find that guidelines that inform data collection can be highly specific and that in simply enumerating the occurrences of certain phenomena the rich array of insights that can arise through field-work may be lost. It is also recognised that visitors are reluctant to criticise exhibits and exhibitions even during in-depth interviews and focus groups and it is often only in visitors' books and comment cards that we find more critical insights and observations. In general, the evidence used to suggest learning or particular forms of learning can appear fragile at best (Culture and Sport Evidence Programme, 2010, p. 59) and little attention is paid to the assessment of the longer-term impact of particular exhibitions.

One further point: there is a methodological conservatism in summative evaluation and we find analytic developments within the social sciences, in particular perhaps those that prioritise the investigation of situated conduct and interaction, largely ignored. We are unsure whether this is due a lack of familiarity with more recent methodological developments in the social sciences or a sense that those that commission evaluation will be reticent to accept findings based on methods that are unfamiliar. The consequence, however, is that developments that might indeed

provide new and distinct findings concerning visitor behaviour and experience remain neglected.

There may well, therefore, be an argument for suggesting that museum evaluation could be enhanced by a more systematic use and awareness of research methods in the social sciences and contemporary developments in data collection and data analysis. An assessment of the 'evidence base' in museums has categorised much of it to in fact be 'consultancy' rather than 'research', arguing that

'good empirical research is grounded, representative and replicable; whereas consultancy is often subjective and may not be representative... Whilst there are a myriad of definitions and interpretations of both of these; it is clear from the review of the evidence base that it contains a mixture of both empirical research (i.e. involving the collection of primary data) and consultancy pieces (i.e. reflecting on personal/organisational opinion). This may not pose any future problems for the sector ... however; if the intention of the evidence base is to provide robust, clear evidence of practice/impacts/outcomes for policy purposes or justifying expenditure then the focus needs to be on research.' (BMG Research, 2009, pp. 21-2)

Similarly,

'case studies and evaluations of specific initiatives [are] frequently successful at telling the story of what was done in any particular initiative, and often capture key outputs and participants' views on what worked. Such work is valuable for the accountability of funding, and can help deliverers and funders to understand how a similar activity might be best carried out and supported in the future. A limitation of such work is that it tends to focus on evaluating one-off, time-limited and pilot schemes. It can be difficult to understand how the findings generalise to day-to-day practice. While there are some examples of excellent evaluations in the evidence reviewed here, there was also a lot of work that is less strong from a research methods perspective. There are studies of initiatives that are conducted before the initiative is completed, and certainly well before any long-term effects could be captured. Many studies do not use before-and-after comparisons to determine what difference a particular initiative made to organisations or participants, and rely on self-reported accounts of the difference made, which makes the findings less reliable. Studies often do not consider the added value of an initiative, in comparison to what would have happened without it.' (Arts Council England, 2011, p. 45)

However, there is an alternative view,

'Evaluation is not the same as academic research. Its purpose is not to increase the sum of human knowledge and understanding but rather to provide practical guidance. If at the end of an evaluation process nothing is changed there was no point in conducting the evaluation. This needs to be the guiding principle in the planning and execution of all evaluation projects.' (Gammon, nd(b), p. 1)

We have some sympathy for this view and, as shown below (p17), observations and findings that derive from research that might be considered methodologically problematic can prove practically useful to those in museums and galleries. However, it is worthwhile reiterating our two principal concerns: (i) that there seems to be limited evidence to suggest that summative evaluation has had a significant impact on practice, and (ii) despite the substantial range of studies that have been undertaken over the past decade or more, they seem to have provided few general findings or widely applicable principles that inform exhibition design and development. Certainly, as in the related area of informal science learning, 'the link between evaluation and research [is] currently underdeveloped' (Wellcome Trust, 2012, p. 5).

Section 2: Concepts and models of the visitor

Using summative evaluation to establish more general conclusions for practice may also be made more difficult because of the concepts and assumptions that inform some evaluations. Critical in this regard are the ways in which some studies conceptualise visitors, their motivation and their behaviour and conduct.

2a) The motivation of visitors

Models of visitor motivation are found particularly, but not solely, in studies undertaken by market research agencies and consultants. Models tend to be particular to a specific research agency and one suspects that they are one way in which an organisation seeks to establish the distinctiveness of its approach and contribution. Since they inform the collection, categorisation and analysis of data, the models contribute to the specificity of the findings of the summative evaluation and can add to the difficulties of comparing and contrasting results between studies that use different models of motivation.

Consider for example Morris Hargreaves McIntyre (MHM), one of the leading consultancies that undertake evaluations for museums and galleries. They segment an 'audience' into particular categories namely sightseers, self-developers, families, repeat social visitors, experts and art-lovers. They explore the relative proportion of each type of visitor to a particular exhibition. (MHM, 2008a, p. 12; MHM, 2008b, p. 15). They also categorise visitors' 'motivations' in terms of four key drivers: Social, Intellectual, Emotional and Spiritual. The motivation categories can be seen as a hierarchy: 'as visitors move up from social through to spiritual motivations they become more increasingly [sic] engaged with the museum and its collections.' (MHM, 2008a, p. 21) Indeed,

'The higher up the hierarchy, the greater the engagement, the more profound the outcome. And, unlike the usual measures of satisfaction, this measure verifies the quality of their experience and even explains how they engaged.' (MHM, 2005, p. 10).

According to one account, the hierarchy of motivation 'reflects Maslow's pyramid of human needs' (Waltl, 2006, p. 5).

No clear or firm evidence appears to be provided for the types of motivation or visitor beyond statements such as,

'The motives listed in this Hierarchy have been painstakingly defined from dozens of focus groups in which visitors reported, articulated and discussed the benefits they were seeking from a visit to a museum or gallery' (MHM, 2005, p. 9).

According to MHM, a successful visit is one in which the visitor's experience is higher in the hierarchy than the visitor's motivation. This in itself is a questionable marker of success: why, precisely, should an emotional experience be superior to an intellectual one?

There are other seemingly idiosyncratic classifications of types of visitor and their motivations. For example, in a recent evaluation, Tate Britain classified visitors into two principal categories namely 'explorers' and 'followers' and despite providing some interesting observations and findings, there seemed to be no evidence or theory that supported the particular classification (Kohler, 2012).

A different classification of visitors has been proposed by Falk. He suggests five: explorer, facilitator, experience-seeker, professional/hobbyist and recharger (previously known as spiritual pilgrim). He argues that a visitor's behaviour can be predicted from his or her 'identity-relative motivation' when entering the gallery. Falk has been criticised as 'too reductive in his treatment of the complexity of visitors' experiences' and for 'the de-emphasising of demographic factors' role in identity construction' (Dawson & Jensen, 2011, pp. 129, 132).

There are a number of difficulties that arise with the use of these various concepts and classification schemes. Firstly, they are rarely mutually exclusive, so the same visitor is likely to fall into more than one category even during a single visit. Secondly, the behaviours and attitudes that evidence the classification of particular visitors into one category or another appear ambiguous and there is little evidence to suggest that they are sensitive to the variety of behaviour that arises during a visit. Thirdly, there seems to be little evidence or theory that supports the selection of the particular concepts used in the evaluation, and at times an uncritical commitment to the idea that behaviour is driven by pre-established personal motivations rather than for example the circumstances that arise during a visit. The problem of motivation and classification is acknowledged. For instance Walmsley suggests,

'people with similar values can be driven by different motivations on different occasions... someone can visit a museum to fulfil several different needs and different people can engage in the same activity for a variety of reasons'
(Walmsley, nd 2011?, p. 1) .

Notwithstanding these difficulties, museums and galleries can find motivational categories illuminating and of practical value. For example, MHM's model has helped the British Museum account for differences in behaviour between visitors to temporary exhibitions and visitors to permanent galleries (Francis, 2011). It has also helped the museum think about different types of behaviour by different visitors, while accepting that an individual's motivations overlap and can change during a visit.

A study of volunteers in the Liverpool City of Culture 2008 makes use of MHM's categories of motivation and observes:

'Without necessarily applying the value judgement of seeing this as a hierarchy, this work does provide a valuable way of categorising different reasons for attendance and participation.' (Rodenhurst & al, 2008, pp. 11-12)
(Interestingly, this report also notes, 'Morris Hargreaves McIntyre suggest that emotional and spiritual engagement, specifically in museums and galleries, is more

effectively experienced without the distraction of others' (Rodenhurst & al, 2008, p. 12), suggesting the MHM model values solitary behaviour over interaction.)

2b) The Social and Interactional

There is a further important methodological issue that informs the character of observations and findings that emerge from evaluation. Notwithstanding significant shifts in how the behaviour and experience of visitors is analysed, to a large extent evaluation has largely focused on the individual visitor and the individual's behaviour and experience. In some cases, even when families or groups are studied, the analytic focus remains primarily concerned with the individual within the collective rather than the ways in which their behaviour and experience of the museum or exhibition arises in and through their interaction with others. This seems at odds with what we know about how people visit museums and galleries and the proportion that visit with others. Indeed, it seems reasonable to assume that the structure of people's visits and how they experience an exhibition arises at least in part through interaction and engagement with others. Even when people visit a museum alone, there are likely to be others within the museum or gallery, and the behaviour of others would appear to have an important impact on navigation, exploration, investigation, and the like (vom Lehn, 2001, Heath, 2004).

It is also widely argued that social interaction is critical to the learning and informal learning that arises in museums and galleries, which makes it more surprising still that the principal focus of many studies remains with the individual visitor, their motivations and the like, rather than how an exhibition or gallery serves to facilitate, even encourage, social interaction and particular forms of conduct, interpretation and experience. There are notable exceptions, such as the Exploratorium's formative Active Prolonged Engagement work (Humphrey, 2005) and a summative evaluation of visitor behaviour at National Trust properties (Boyd & Graham, 2010).

2c) Where the action is

Perhaps more surprising still, is that summative evaluations largely disregard what people say and do at the 'exhibit face' - when they are actually looking at exhibits. Methods such as surveys, in-depth interviews and focus groups, elicit information from visitors concerning their behaviour and experience, but pay limited attention to the actual conduct of visitors as it arises when they are looking at and discussing exhibits. The complexities, details and contingencies of visitors' actions and activities, their talk and visible conduct that arises during the course of a visit remain largely neglected and yet one might imagine that what people think, say and do *in situ* lies at the very heart of the museum visit and the individual's experience.

Visitor engagement with individual exhibits is perhaps more the preserve of formative evaluation and there is a particularly strong tradition of prototyping interactive science exhibits and testing them with visitors, making incremental improvements (Gammon & al, 2011) (Humphrey, 2005). However, this is far less common in most other areas. Moreover, the fact that an exhibit has been subject to formative evaluation does not justify neglecting how visitors engage with the exhibit on the gallery floor, as visitor behaviour can be strongly influenced by the context of

the exhibit.¹ In one very thorough evaluation it is clear that even though interactive science exhibits were subject to exhaustive formative evaluation, many visitors did not engage successfully with them *in situ* (Gammon & al, 2011).

The growing commitment to undertaking observation as part of evaluation is to be welcomed. It is widely recognised, however, the participant and non-participant observation, even undertaken by the most observant researcher, provides limited access to the details of the participants' conduct, the talk and bodily comportment that arises as people navigate museum spaces, explore and examine exhibits.

It is perhaps surprising, given the wide-spread availability of a cheap and reliable technology that enables you to capture both the spoken and visible conduct of visitors *in situ*, that video has not had a more important impact on summative evaluation (although, again, there are notable examples - for example in formative evaluation (Humphrey, 2005)). Video provides unprecedented access to human behaviour, enabling researchers to capture versions of conduct and interaction as it arises on particular occasions and subject it to repeated scrutiny. It also enables researchers to examine and share the raw data with others, including museum personnel and visitors and invite their comments and reflections on the actions and activities that are found in the recordings. Used more widely, video recordings could provide a raw data corpus that would enable comparative analysis between different exhibitions, museums and galleries, yet it is rarely used in evaluation. Indeed, it reminds one of Margaret Mead's comments concerning the absence of the use of film in social anthropology in the 1960s and 1970s,

...research project after research project fail to include filming and insist on continuing the hopelessly inadequate note-taking of an earlier age, while the behaviour that film could have caught and preserved for centuries ... (... for illumination of future generations of human scientists) disappears – disappears right in front of everybody's eyes. Why? What has gone wrong? (Mead, 1995, pp. 4-5)

Summary of Sections 1 and 2: Methodological and conceptual issues

In summary, therefore, there are a number of methodological and conceptual problems and issues that undermine drawing more general conclusions from summative evaluations and using this wide range of studies to build a corpus of findings and recommendations concerning the engagement of visitors and development of museum and gallery spaces. These include:

Data Collection and analysis

- the diverse range and combination of methods used in different evaluations to collect and analyse data

¹ To address this, the Exploratorium's admirable Active Prolonged Engagement research intended that the exhibits subject to visitor research 'would be distributed among other exhibits'. However, in practice 'we created a special videotaping area cordoned off from the rest of the museum floor...we posted signs at the entrances informing visitors of the research and of the videotaping... we also used sound-absorbing padding on the walls, ceiling and floor', so the setting was rather different to a normal visitor experience (Humphrey, 2005, p. 3).

- the varied ways in which data is analysed and presented
- the narrow focus of summative evaluation structured to address the specific aims and objectives of particular projects
- the methodological shortcomings and conservatism in the use and application of particular techniques.

Concepts and models of the visitor

- different models and conceptions of visitors and their behaviour that underpin different forms of evaluation
- the relative neglect of the importance of social interaction to visitors' experiences of museums and gallery
- the disregard for the *in situ* details of action and activity that arise at the 'exhibit face'.

Section 3: The organisational and institutional framework

Methodological problems and issues alone do not explain the relative lack of impact of summative evaluation - indeed it is not the principal problem. One of the most important areas to consider is the organisational and institutional framework in which major projects are commissioned, undertaken and evaluated. There is neither the space nor remit here to address these matters in detail, but it is useful to consider several issues and practical problems.

3a) Specific objectives, not general questions

Major initiatives in museums and galleries frequently involve raising significant funding from external organisations such as HLF, Wellcome, or other trusts and foundations. The expectations of these external organisations reflect the contemporary criteria that inform new displays - for example, the emphasis on enhancing access, participation and learning. Many summative evaluations are produced not only for internal consumption but also (if not primarily) for the funders (Griggs, 2012a). The summative evaluation is understandably tailored with regard to the aims, objectives and specification of the original project and the overall commitments reflected in the funding. Its principal concern is to assess the extent to which the project met its aims and objectives and not to produce more general observations and findings: the success of meeting specific objectives, more than wider outputs and outcomes (Doeser, 2012). Therefore, it can be too inward-looking and focused to be generalizable (Griggs, 2012b). This makes sense in terms of assessing the specific project, but undermines the broader relevance of the research.

However, it does not make it impossible. As one evaluator commented,

‘A lot of the evaluation of permanent galleries and temporary exhibitions I’ve been involved in has been quite specific but there are certainly findings from many of these that could easily be transferred to other projects within the organisation or to similar organisations.’ (Boyd, 2012a)

The potential for subsequent applicability within an individual museum will depend in part on the nature of the project subject to summative evaluation. It will naturally be easier to apply findings if something similar takes place. So, for example, when designing the reading room for each year’s Turner Prize exhibition at Tate Britain it is possible to quite straightforwardly learn from evaluations of earlier reading rooms (Koutsika, 2012). In contrast, with its varied programme, it may be less straightforward for the Wellcome Collection to draw lessons from the evaluation of one exhibition and apply them to a subsequent very different one.

3b) Monitoring and accountability - and the role of the funder

Summative evaluations that are produced in order to satisfy funders effectively form part of the funder’s monitoring and accountability process. This immediately constrains the evaluation; indeed, some advice recommends that evaluation is

clearly distinguished from monitoring, 'there is often confusion between monitoring and evaluation.' (RCUK, 2011, p. 2) Funders regularly say that they see the primary purpose of evaluation as helping people reflect, and want to see people learning from evaluation, but concede that many of the summative evaluations that they receive tend to have a positive bias. For example, the Wellcome Trust wants organisations to learn from their mistakes as well as their successes and endeavours to explain that as long as that learning takes place, honest reporting of difficulties, or even failures, will not jeopardise future funding applications. However, organisations still often appear reluctant to submit frank evaluation reports (Stewart, 2012).

A summative evaluation is often seen as the closure report on a finished project, designed primarily to show that project objectives have been achieved (Pegram, 2012a). Even though HLF provides extensive advice on evaluation, with the help of the New Economics Foundation (Heritage Lottery Fund, 2008), some of the summative evaluations it receives are poor quality - indistinguishable from end-of-project reports. This might be exacerbated by the fact that HLF often requires the submission of an evaluation report to trigger payment of the final stage of a grant – creating an expectation that HLF wants proof of success, whereas HLF would prefer an honest and genuine account of what happened, with the evaluation acting as a resource from which the museum can learn (Talbot, 2012). However, 'Essentially, the evaluation report has been prepared to tick somebody else's boxes – a funder usually – and the opportunity to use it to reflect upon and learn from experience is lost. Instead, it gets quietly filed with all the other reports, never to be seen again.' (Forrest, 2012)

The role of the funder is difficult to get right. There are arguments that funders should be firmer in specifying what they require from evaluation, for example specifying how evaluation should be approached in detail to ensure consistency of definition in areas such as demographics and to achieve better comparability and even economies of scale (Groves, 2012). However, funders don't want to be seen as excessively prescriptive and bureaucratic; they prefer that individual organisations should determine their own most suitable approach and intend to allow space for variety and creativity (Talbot, 2012). Nevertheless, summative evaluation is often undertaken as a consequence of the requirements of the funding received by the museum or gallery to undertake a particular project. We believe that there are significant benefits in funders taking more responsibility for and playing a more active role in evaluation. For example, it might be helpful for some funders to revisit their expectation of front-end evaluation at the start of a project and summative at the end in favour of a more continuous process (Koutsika, 2012), so that more evaluation takes place at a time when it is still possible to make changes relatively easily (MacDonald, 2012). Furthermore, linking summative evaluation to final grant payment can force it to be done to tight timescales, which can restrict the techniques used, put the focus on short-term impacts and lead the gallery to be prematurely regarded as a finished product (Moussouri, 2012). (During the course of our investigation, several people commented that it would be useful to have evaluation of museums' longer-term effects on audiences. A recent review of informal science learning similarly noted the importance of understanding long-term

impacts and observed: ‘Considerable effort, and investment, is directed towards evaluation across the sector, although for the most part attention focuses on process issues and capturing immediate or short-term outcomes. The assessment of longer-term impacts is rare.’ (GHK Consulting, 2012, p. 37))

Some further suggestions on the role of funders are included below in 3d) and in section 5.

3c) Celebration and advocacy

There is a risk that summative evaluations can become celebratory. Indeed, museums (and funders) can talk of evaluation being helpful to ‘prove’ or ‘demonstrate’ a project’s success. Furthermore, museums and galleries will often want to use the findings of an evaluation study for future advocacy - to demonstrate their past successes in order to build future support. In many cases this leads to an expectation that the evaluation will identify the positive contributions of the project. It is unusual (but not unknown) to find examples of evaluation that significantly criticise the particular initiative.

This is not to suggest that those undertaking the evaluation are biased, but rather the requirement for good news forms the framework in which evaluation takes place. One evaluator explained:

‘I’m currently working on an evaluation that requires a final report “in the form of an advocacy document”. This isn’t evaluation and I’ve had that conversation with the client but this says a lot, not only about their attitude towards evaluation but also about where the sector is at the moment – frightened and desperate to shore up its position with policymakers and potential funders.’ (King, 2012a)

On occasion, it appears that different versions of evaluations are produced for circulation beyond the project team or beyond the organisation, ‘evaluations are edited down to play down problems’ (King, 2012a). In consequence, mistakes and difficulties are rarely brought to fore in published evaluations, even if they were made in the original ‘internal’ version. There are some exceptions to this, such as the published highly critical summative evaluation of the V&A’s Jameel Gallery (Fusion, 2008).

3d) Fear of exposure

Many of the individuals directly involved in a project do want to receive an evaluation that points out difficulties and shortcomings. The individuals can learn from their mistakes – and in principle could correct some of them, although there are practical constraints on making remedial changes (see 3g).

Individuals and project teams need to be self-confident if they are to be willing to share more widely information about weaknesses. Sometimes museums say they want a ‘warts and all’ evaluation, but change their stance on receiving one (King, 2012b). In some cases there is a reluctance to share ‘warts and all’ evaluation, even within a museum. A participant in the first Evaluating Evaluation colloquium was

concerned about ‘evaluation being used as an internal “weapon” rather than for shared learning’. Even if the project team, or audience research staff, want to share the findings of evaluation, it might not be valued by some of those they wish to share it with. Participants in the first colloquium said ‘many do not recognise potential value of evaluation; some may fear it’; and some may even have ‘fear of being audience-led’.

Some organisations publish their summative (and other) evaluations, but others appear reluctant to share the findings of evaluation with the wider sector. Fears of damage to reputation, and a sense of competition between organisations, make widespread sharing of honest evaluations difficult to achieve, particularly if the museum has not acted on the findings. There is a risk that a sponsor will be alarmed by a starkly worded critical evaluation.

The reticence to share may have been exacerbated in recent years by the financial pressures on museums. Some individuals genuinely fear for the continued existence of their jobs and there is increasing competition between museums for external funds.

Funders will need to work hard to reassure museums and convince them of the benefits if they genuinely want honest evaluation. But it may be that funders do not in fact want evaluation that reports fully on failure. Within funding organisations, staff and departments can be in competition for funding for their area of work and may fear sharing bad news with senior staff and board members who allocate resources. Some funders have external accountabilities, for example to government, and may fear damage to their reputation. To an extent, a poor project can reflect badly on the funder, for making the decision to support the project in the first place.

Having said that, within some museums (and funders) there appears to be quite extensive reflection and learning on the findings of evaluation within groups of staff – and sometimes more widely within the organisation. In addition, there are networks of practitioners in different organisations that informally – and sometimes confidentially - share evaluation results between them. On the other hand, some independent evaluation consultants are in competition with each other, and so reluctant to share information.

Case Study 1 The Wellcome Trust

The Wellcome Trust’s Engaging Science grants team see two main purposes for evaluation. Firstly, in helping those running a project to understand whether they have done what they set out to do (ideally this means projects need clear objectives at their outset); and secondly, to provide accountability to funders for what their money has been spent on. They want grantholders to learn from their evaluations and to explain how they will apply what they have learnt to future projects, as well as sharing any learning with the wider public engagement community.

If an applicant can demonstrate they have learnt from what has happened in their project, both positive and negative, it can increase their chance of securing funding

for future projects (although it is still primarily dependent on the quality of their new project idea). The grants team want 'warts and all' evaluation reports, with grantholders being honest about what worked and what didn't. They read all the evaluation reports they receive, try to visit all the projects they fund and then complete final assessments of projects as part of the internal accountability to their board of governors on how their funding is spent.

Wellcome has developed a framework to assess outcomes that will help them rate the impact, reach, quality and value for money of projects consistently; they are also developing a regular series of meetings with grantholders to help them understand this assessment framework and what it means for their reporting requirements.

Wellcome may move to a position of 'strongly encouraging' grant holders to publish evaluations so that others can learn from them. However, they accept that a summary of key learning points may be more appropriate for this purpose than publishing complete evaluation reports.

3e) The fragmentation of knowledge and expertise

It is not uncommon for a team to be assembled for the duration of a particular project and then disbanded following the completion of the gallery or exhibition. The team may well include a temporary assembly of internal staff of the particular institution – the relevant curator(s), subject experts, educationalists, interpreters, designers, project manager - and in some cases, members of an external design company that has secured the commission. Some staff may be on short-term contracts that last only for the duration of the development of the gallery or exhibition. The team develops extensive expertise over the course of a project concerning the design of displays and exhibits, the provision of interpretative resources, and an understanding of visitors and their response. Unfortunately, in many cases, there no formal or even informal mechanism through which that expertise and knowledge is preserved and made accessible to those involved in subsequent projects. To make matters worse, it is not uncommon for project teams to be dispersed prior to the completion of evaluation, or its review and application, so expertise and critical reflection can be lost. As King (2012a) suggests, 'Project teams disperse before learning can be acted upon and the next project team rarely starts by analysing what they can learn from the last one.' Moussouri (2012) suggests, conversely, that when findings of earlier evaluations are discussed by a project team, the resulting decisions may not be recognised as being a result of learning from evaluation but may be erroneously remembered as ideas coming from the project team itself. Also, fragmentation of teams does not in fact mean destruction of expertise: there are people, such as former Science Museum staff, throughout the sector who draw regularly on their tacit knowledge of visitor behaviour which in many cases originally came from evaluation studies.

We have also found that summative evaluations are rarely made available to the external design consultants involved in a particular project, so an important opportunity to transfer and preserve knowledge across particular cases and

institutions remains unexploited. Moreover, design companies develop a substantial body of working knowledge and practice that is applied through successive projects and although this tacit knowledge is applied by the companies to their subsequent projects, it remains largely unknown and inaccessible to museum personnel.

Some museums, such as the British Museum – see below - have internal mechanisms to retain learning about the efficacy of displays and exhibitions. Other museums have brought together the findings of a series of evaluations and studies of visitors to attempt to draw general lessons (Pontin, 2010). ‘Sometimes I just think the evaluation/consultation data needs a little re-packaging to make it useful.’ (Boyd, 2012a) As an external evaluator, Nicky Boyd has strategies to encourage museums to take note of her evaluation reports – she tries to organise a meeting or workshop at which she helps staff identify things they will do as a result of the evaluation’s findings, and she likes to ask museum clients to write a preface to the evaluation report (Boyd, 2012b).

Case Study 2 The British Museum

Since 2005 the British Museum has undertaken systematic summative evaluation. Evaluation is overseen by the interpretation team in collaboration with marketing and exhibitions. Summative evaluation of smaller more experimental exhibitions has proved to be particularly informative and influential. The museum now has an extensive archive of data on visitor behaviour in permanent galleries and temporary exhibitions. Summative evaluation is seen as a way of building knowledge about visitor understanding and behaviour that can then be implemented in future displays. Exhibition evaluations enable comparisons to be made between displays, but they also allow the effectiveness of elements specific to an exhibition – such as an experimental approach to display or the use of a new piece of digital media- to be assessed. The cumulative knowledge of visitor behaviour is used to help the museum become more effective at increasing visitors’ depth of engagement.

In planning new displays and exhibitions, evaluation is used to help the project team make informed decisions on the basis of evidence from research rather than the opinions or experiences of staff – ‘folk knowledge’ does not always reflect reality. Evaluation can influence team members, strengthen arguments and depersonalise disagreements. Staff working on exhibitions at the British Museum have learned that in spite of learning from evaluations, and applying the lessons to future projects, unexpected outcomes still arise. It is important to be open-minded and willing to acknowledge failures as well as successes.

Indicators of visitors’ depth of engagement in different permanent galleries, whether the average dwell time or the average number of stops made, can be compared. A ‘best fit’ line is used to help compare dwell time in galleries of different size. In the case of temporary exhibitions, the museum is aiming to use its evaluation archive to establish helpful parameters such as the optimum number of sections, objects or words to create a manageable visit (for example 75 minutes).

ctd .../...

The museum evaluates permanent galleries before they close for refurbishment both to understand visitor behaviour, to see what works well and what doesn't, and also to give a baseline against which to judge the success of the replacement gallery.

Summative evaluation has demonstrated most visitors to older permanent galleries at the BM do not read introductory panels or follow the intellectual framework systematically. Nevertheless visitors still expect a gallery to have a clear structure and coherent organising principles, otherwise they feel disorientated. Evaluation revealed that visitors in the permanent galleries tend to start with an object, rather than a panel or text, This has led the museum to develop gateway objects as an alternative to panels introducing each section or subject (Francis, 2011). Single prominent objects are integrated with layered texts which introduce a section or showcase by starting from the object and working outwards. Summative evaluation indicates the approach is effective. Typically, a refurbished gallery with gateway objects has twice the average dwell time of the gallery it replaced or of older galleries which retain traditional hierarchies.

BM evaluations are undertaken by external consultants, members of staff and students on placement with the interpretation team. A number of evaluations that have resulted in formal reports of publishable quality are available on the BM website.

3f) The position of the evaluator

Much evaluation is undertaken by external consultants. This can bring an impartial view and a useful sense of distance. However, several independent evaluators contacted us to point out disadvantages². These include problematic briefs prepared by museum clients that might expect far more than is reasonably achievable for the fee being offered, or be unclear about precisely what is to be evaluated, perhaps because the museum did not set clear objectives for the display or exhibition. Project objectives can be 'eye-wateringly' over ambitious because the project over-promised what it could achieve. It can be helpful to involve evaluators early on in a project so they can help set clear, measurable objectives amenable to later evaluation (Wilkins, 2012). In some museums staff may lack confidence in briefing, selecting and commissioning evaluators (King, 2012b).

Sometimes external evaluators feel that the client acts as a 'gatekeeper', restricting the evaluator's access to people and information. One evaluator gave the example of 'being kept apart from other key players, particularly designers'. Sometimes, external evaluators can feel marginalised, brought in only because of the requirements of the funder. In-house evaluators sometimes say they feel marginalised, too, and it is clear that individual museums have very varied attitudes to evaluation. A recent study of the relationship between internal museum cultures and the impact of audience research concluded that the key factors that influenced whether audience research (including evaluation) had an impact on a museum's

² The evaluators concerned asked to remain anonymous

practice are ‘integration, acceptance, senior management support, audience orientation and utility of the studies.’ That is, ‘The effectiveness of audience research therefore does not rely on the person conducting the research alone but requires the involvement, commitment and support of senior management and museum staff.’ (Reussner, 2009, p. 446)

Moreover, external consultants are in a potentially vulnerable position. If an evaluation is overly critical they can be fearful that the organisation is unlikely to commission further research from them.

3g) The impracticality of remedial change

There is a flaw at the heart of summative evaluation: it usually proves impractical to make significant changes to a permanent display or exhibition once it is complete. Unlike formative evaluation, by the time summative evaluation has been undertaken it is usually too late to make changes, particularly because resources are not usually available for remedial work. It is also psychologically more difficult to change things once they are completed, rather than at an earlier stage (Griggs, 2012a). There are some examples of small changes being made in response to evaluation, for example to signage or repositioning popular objects to make them easier for visitors to see. However, it appears even small changes can take a long time to be implemented. Many displays and exhibitions are complex and costly to alter and money is rarely held back to cover the cost of changes (Higgins, 2012b).

To support and encourage small changes, it is in principle possible to break summative evaluation into two stages. The first ‘monitoring’ can look rapidly at how visitors use a gallery and inform quick changes; later a fuller ‘verdict’ evaluation can more fully assess the gallery (Wilkins, 2012).

Summary of Section 3: Organisational and Institutional Framework

The organisational and institutional environment in which evaluation is commissioned, undertaken and disseminated can undermine the impact of the findings and the development of a body of shared knowledge and expertise. The constraints placed on summative evaluation include:

- the ways in which studies are tailored with regard to the specific requirements of particular initiatives
- the necessity to provide a positive assessment of the project’s outcomes and the achievement of its goals
- the lack of opportunity or mechanisms to share experience and establish a repository of knowledge and practice both within and across institutions
- the limits placed on access to reports outside, and in some cases, within particular organisations
- the temporary, project specific, organisation of design and development teams especially in the case of major initiatives
- the use of external design consultants who temporarily bring their expertise and knowledge to bear on the development of exhibitions and exhibits but who are not included in the evaluation process

- the detached or marginal institutional position of some of those individuals or organisations who undertake evaluation
- the impracticality and expense of undertaking 'remedial' changes to an exhibition, gallery or even exhibit

Section 4: Conflicting purposes

It is clear from the above discussion of the frameworks surrounding summative evaluation that it can be subject to different purposes, particularly:

- reflection on the particular project and learning within the project team
- learning and change within the wider organisation
- knowledge and learning across the sector
- monitoring and accountability to both the institution and the funder
- advocacy, for example in support of bids for future funding

There are conflicts between these purposes. For example, if an evaluation report is to stimulate frank reflection within the project team, it is unlikely to be suitable as a document for advocacy, and vice versa. As discussed in 3d, people involved in a project can be reluctant to share findings of problems more widely within the museum, let alone with the wider sector or with funders. As discussed in 3b, people are understandably reluctant to share everything with funders for fear of jeopardising future applications to the funder.

Many of the problems with summative evaluation arise from these conflicting purposes. Requiring summative evaluation has now become the norm but we are not convinced that funders or organisations are always clear about what they want when they request a summative evaluation. Those requiring or commissioning summative evaluation should perhaps first clarify *why* they want the summative evaluation and what they intend it to achieve. Museums need 'clarity of what summative evaluation is there to do; what is really useful for us to know – and use resources to do that' (Pegram, 2012b).

Section 5: Increasing the impact of summative evaluation: Key recommendations

In the course of our investigation we found examples of good practice in which summative evaluations within particular institutions systematically inform subsequent work. In addition, a range of people suggested to us how the impact of evaluation, both within the host institution and more generally, might be enhanced. Here, we draw on these observations and suggestions to make recommendations for improving the impact of summative evaluation, depending on the purpose of the particular piece of summative evaluation. We conclude with a series of recommendations addressed to museums, to evaluators and to funders. Fuller details of suggestions for improving the impact of summative evaluation can be found in appendix 1.

Summative evaluation will be more effective if there is clarity about its purpose. The first task of a funder or museum requiring summative evaluation is to specify what it is intended to achieve and how it will be used. Expectations of summative evaluations will vary, as there will be different drivers and contexts for each particular piece of evaluation. There may be external drivers (particularly funders) and internal drivers (such as whether the project is part of a series or a one-off); there may be personal drivers coming from the interests and experience of the people commissioning the evaluation (Koutsika, 2012). Briefing is key; museums' evaluation briefs should be clear about the use(s) to which the evaluation will be put, how it will relate to other evaluation and audience research, and how it will be disseminated. Briefs should be realistic and not set too many objectives or requirements for the resources available.

Depending on the aim of a summative evaluation, different actions will help increase its impact.

A Aim: Encouraging reflection by the project team

There are a number of ways of increasing the impact of summative evaluation on reflection and learning by the project team. Most of these are very simple and straightforward and will cost little apart from staff time. Full details are given in appendix 1; key points include:

- Think about summative evaluation early on in the project
- Include representation of evaluation (or audience research/visitor studies) on the project board
- Ensure the project team has enough time to reflect on (and ideally act on) the findings of evaluation

B Aim: Supporting the impact of evaluation within the wider museum

Possibilities here fall into four main areas (again, for fuller details see appendix 1):

- Embed evaluation in the museum

- Have mechanisms and structures for raising awareness of evaluation, retaining knowledge from it and demonstrating its relevance to a range of staff
- Have an individual or department with clear responsibility for evaluation and building up and applying knowledge long-term
- Build in time and money for remedial change, to act on the findings of an evaluation

C Aim: Supporting knowledge and learning across the sector more broadly

Potential ways of improving the impact of summative evaluation more broadly include:

- Do more to share the findings of evaluations
- Consider the merits of publishing all summative evaluations, or at least key findings
- It would be helpful to produce be syntheses of findings from a range of evaluation studies to identify common points. For this to be effective, evaluations would need to be designed to support synthesis
- Identify general research questions to give a wider context for summative evaluation
- Raise awareness of the potential benefits of evaluation

Again, more detailed suggestions can be found in appendix 1.

Case Study 3 Networks for sharing evaluations

There are a number of networks for sharing evaluation. Probably the most extensive is www.informalscience.org, 'a resource and online community for informal learning projects, research and evaluation' which is administered by the University of Pittsburgh's Center for Learning in Out-of-School Environments and funded by the National Science Foundation. In July 2012 it included over 400 full texts of evaluations, of which around 150 cover museums, and abstracts of over 5000 pieces of research.

In the UK <http://collectivememory.britishtscienceassociation.org/> is 'a database of evaluations of a diverse range of science communication initiatives', developed by the Science in Society team of the British Science Association supported by the Department for Business, Innovation & Skills 'Science for All' expert group. It encourages informal science learning projects to post brief details, including summaries of 'what went well', 'what was learned' and 'hints and tips for others'. The entries do not include full-text evaluations and appear to mainly cover activity projects rather than exhibits.

The Visitor Studies Group <http://www.visitors.org.uk/> acts as a network for people involved in visitor studies through events and an e-mail list and has a small selection of evaluations on its website.

D Aim: Monitoring and accountability

Here, summative evaluation as currently undertaken can have a role; as one anonymous participant in the second Evaluating Evaluation colloquium explained, ‘Summative evaluation sometimes has a very specific remit... Without summative evaluation we wouldn’t know for sure whether we’d been successful or not. We wouldn’t know exactly who had come etc. We can’t rely on what we think about an exhibition to measure its success.’ However, as noted above, museums can be reluctant to report problems to funders, and evaluation can become celebratory, so it may be unwise for funders to see museum-commissioned evaluation as a reliable source of monitoring and accountability. As a quality control and audit, funders are advised to directly commission (and publish) fully independent evaluation of a sample of projects.

Funders generally have separate monitoring processes and for good evaluation it is important to distinguish between evaluation and processes for monitoring or accountability.

E Aim: Advocacy

Certainly, summative evaluation can provide examples of successful achievement by museums, but done well it will also find evidence of less successful outcomes, so it is perhaps best not seen as directly contributing to a museum’s advocacy work.

Summative evaluation may not be the answer

It may be that in some cases summative evaluation is not in fact the solution to the problem the organisation wants to address. If the project team wants help with reflection then it may in fact be better to implement a continual process of evaluation than runs closely alongside the project; if wider learning is required then it might be better to commission something closer to research; if it is monitoring and accountability, then something closer to an audit or inspection might be most suitable; if it is advocacy that is required, then the museum might in fact want a public-relations expert rather than an evaluator.

Key recommendations

We would like to draw particular attention to the following recommendations to museums, to evaluators and to funders. We hope all three groups, perhaps working together, will find ways to respond to these recommendations. Perhaps a starting point could be a concerted push by the evaluation community (that is evaluators themselves and funders and museum staff who recognise the benefits of evaluation) to demonstrate the value of evaluation.

A) Recommendations for Museums

A1 The first task of a museum requiring summative evaluation is to specify what it is intended to achieve and how it will be used. Evaluation briefs should be clear about the use(s) to which the evaluation will be put, how it will relate to other evaluation and audience research, and how it will be disseminated within the museum and possibly more widely.

A2 Embed evaluation in the museum. Have an evaluation framework and include evaluation as a key part of other plans and strategies. Allocate responsibilities for evaluation and audience research. Have mechanisms in place for preserving and disseminating knowledge from evaluations among staff and across successive projects.

A3 Disseminate the findings and implications of summative evaluations across the broader community including museum staff, evaluation consultants and designers, by publishing reports, or digests of key findings, and organising or contributing to workshops, seminars and conferences.

A4 Identify overarching research questions to inform individual pieces of summative evaluation. Consider collaborating with other museums to establish a common set of research questions and secure comparable data.

A5 Enable and support comparative analysis and the synthesis of findings from a range of summative evaluation studies through adopting at least some common themes, adopting similar research methods and building an accessible data corpus and comparable analytic framework. Consider raising funding, or building partnerships, to enable this to happen.

A6 Plan for change. Build in time and money for remedial change, to act on the findings of an evaluation.

b) Recommendations for Evaluators

Some of these recommendations apply to individual evaluators; others to evaluators working together, possibly via an organisation such as the Visitor Studies Group

B1 Raise awareness of the benefits of evaluation across the sector. Explore ways of increasing museum practitioners' skills in commissioning and understanding evaluation. Training or good-practice advice on briefing could be particularly effective. Showcase examples of best practice, including examples where evaluation has had a significant impact on practice.

B2 Share the findings of evaluations. When possible, publish reports, or at least digests of key findings. Contribute to seminars and conferences.

B3 Enable and support the comparison and synthesis of findings from a range of summative evaluation studies. Consider raising funding, or building partnerships, to enable this to happen.

B4 Devise overarching research questions and encourage museums to support summative evaluations that contribute to the research questions.

B5 Consider standardising some aspects of data collection and methodology to aid comparability of evaluation studies.

B6 Exploit contemporary methodological developments in the social sciences and consider using more innovative ways of collecting and analysing data.

c) Recommendations for Funders

C1 Be precise in specifying what summative evaluation is intended to achieve.

C2 Engage with grant-recipients' evaluation work: set high expectations of transparency and honesty; ask questions of the report.

C3 Encourage (or require) museums to act on the findings of summative evaluation and provide funding to allow changes to be made to galleries. Consider requiring plans for acting on the findings of each evaluation.

C4 Raise awareness of the potential benefits of evaluation across the sector. Explore ways of increasing museum practitioners' skills in commissioning and understanding evaluation. Training or good-practice advice on briefing could be particularly effective.

C5 Encourage dissemination and sharing of the findings of summative evaluation studies, and synthesising the findings of a range of evaluation studies. Consider taking responsibility for ensuring these activities happen.

C6 As a quality control and audit, directly commission (and publish) fully independent evaluation of a sample of projects.

Section 6: Overall conclusion

Overall, after a year of reading, talking about and analysing summative evaluation, our feeling is probably best expressed as one of disappointment. Disappointment that all the energy and effort that has been put into summative evaluation appears to have had so little overall impact, and disappointment that so many evaluations say little that is *useful*, as opposed to merely interesting.

There are, of course, exceptions. There are some museums where evaluation has a significant impact on practice and there are some evaluations that begin to get to the heart of understanding the visitor experience – telling museums what works and what could be improved.

At the start of our investigation, we felt that perhaps the key problem with summative evaluation was methodological and conceptual weaknesses, as outlined in sections 1 and 2. However, as our work progressed we came to believe that increasing the rigour of summative evaluation would not, by itself, make a huge difference. A more fundamental problem is the organisational and institutional framework in which summative evaluation takes place, as discussed in section 3. This belief was reinforced when we heard from several leading evaluation consultants, who passionately expressed deep frustration about the frameworks within which they have to work.

Overall, and with some exceptions, summative evaluation is not taken seriously enough or well enough understood by museums, policy makers and funders. The visibility of summative evaluation is low. Too often, it is seen as a necessary chore, part of accountability but marginal to the core work of museums.

Perhaps a starting point to change this could be a concerted push by the evaluation community (that is evaluators themselves and funders and museum staff who recognise the benefits of evaluation) to demonstrate the value of evaluation. Museums that benefit from evaluation could do more to raise its profile internally and externally. They could at the least publish evaluation reports, with commentary on how they are talking the findings into account. They could disseminate the key findings more widely through articles, conference presentations and seminars to explore the implications of individual evaluation studies. This would show practitioners how they could benefit from evaluation and would also help develop a stronger culture of criticism of permanent galleries, which are rarely reviewed either publically or professionally. (There are many more reviews of temporary exhibitions and entire new or refurbished museums.) This could then influence strategic bodies, policy makers and other funders to place more importance on learning from summative evaluation.

We are pleased that the Visitor Studies Group is intending to play a role in taking forward the findings of the Evaluating Evaluation investigation, for example by organising seminars and discussions.

Museums in England are increasingly being encouraged to strive for 'excellence' and this may give an opportunity to enhance the role of evaluation, which, done well, can show what makes excellent, engaging visitor experiences.

However, it's worth noting that some museums may still fail to learn from the experience of others, people on a project team 'want to go on adventure together and not hear people say "evidence shows that this won't work"... This means people don't learn from evidence and make the same mistakes, together, over again. Clients can be resistant to hearing advice about what worked and what didn't in another museum.' (Casson, 2012)

Furthermore, while the head says that work should be evidence-based, some of the best projects come from the heart or the gut (Featherstone, 2012). Additionally, the behaviour of visitors can be maddeningly (or thrillingly) unpredictable (Arnold, 2012). To a degree, museum work may be more of a craft than a profession. Perhaps, in fact, 'practitioners can be characterised more as craftspeople operating through a model of apprenticeship, observation and audience approval. This contrasts with the 'professional' tradition whereby formalised mechanisms are developed to ensure knowledge is recorded and training is made available to both new and existing entrants to the field.' (Wellcome Trust, 2012, p. 5)

Acknowledgments

Thanks to the Wellcome Trust and Heritage Lottery Fund for funding this research. Thanks in particular to Ken Arnold and Ben Stewart at Wellcome and to Karen Brookfield, Gareth Maer and Fiona Talbott at HLF. Thanks to all the people who spoke at the two colloquia and to everyone who attended and made a contribution. Thanks, too, to people who made themselves available for interview and who sent us thoughts and comments.

Appendix 1

Suggestions for improving the impact of summative evaluation

In the course of our investigation many people have suggested how the impact of evaluation, both within the host institution and more generally, might be enhanced. We also found examples of good practice in which summative evaluations within particular institutions systematically inform subsequent work. Here, we draw on these suggestions and observations to suggest how the impact of summative evaluation can be improved, depending on the purpose of the particular piece of summative evaluation.

To increase impact on reflection by the project team

There are a number of ways of increasing the impact of summative evaluation on reflection and learning by the project team. Most if these are very simple and straightforward (although some will require staff time)

- Think about summative evaluation early on in the project, involving the whole project team in discussions, rather than trying to design it at the end; this will also help ensure that the project has reasonable objectives
- Evaluators could do more to raise awareness of the varied possibilities of evaluation, so museums don't simply repeat what they have always done (Wilkins, 2012). More creative and unusual forms of evaluation might come up with more unexpected and interesting findings and there are benefits in peer review (Arnold, 2012), although some fear peer review risks marginalising the views of the audience (Featherstone, 2012)
- Make sure evaluation (or audience research/visitor studies) is represented on the project board
- Make sure the brief includes:
 - A clear role for the evaluator to help the project team learn from the evaluation. This may include contribution by the evaluator to a process of reflection by the project team, after the evaluation is complete. One approach is to run a workshop that will aim to identify, say, five key learning points from the evaluation and then to ask a member of the project team (or the museum as a whole) to write a preface to the evaluation report (Boyd, 2012b)
 - A requirement for the evaluator to contact all members of the project team (or in a large team, representatives of all members)
 - A requirement for the evaluation to gather data and undertake analysis that will provide at least some comparative findings with other evaluations undertaken by the organisation
 - Clear identification of the aspects of the gallery/exhibition that were experimental and so require specific attention
- Have, in the words of a contributor to the final colloquium, a 'range of post-project meetings that focus constructively on learning from the experience.'

Make sure the project team has enough time to analyse and reflect on (and ideally act on) the findings of evaluation. In some cases this may necessitate reconvening the project team to include external consultants, such as designers, or people who were on now-complete fixed-term contracts

To increase impact within the wider museum

Possibilities here fall into four main areas:

i) Embed evaluation in the museum. Possibilities include:

- Include evaluation/visitor studies on every appropriate project team. In some museums evaluation is the responsibility of the interpretation team and is therefore a consistent element in every redisplay or exhibition team (see BM case study)
- Have a clear framework or strategy for evaluation (that also includes, or at least references, market-research undertaken by the organisation)
- Champion evaluation from a high level in the organisation, incorporate into strategy and planning, such as annual business plans
- Create a 'culture of reflective/learning organisation that learns from mistakes and from taking risks' (Koutsika, 2012)
- Have, and include evaluation in, audience development and learning strategies
- As each organisation has a different internal culture, external evaluation consultants need to try to understand how a particular organisation works and devise strategies to help their work have the greatest impact (Boyd, 2012b)
- Don't only evaluate projects; also evaluate more day-to-day work against longer-term plans and outcomes – to do this need a conceptual structure for what the museum is trying to achieve overall, so projects and their evaluation are seen as part of a continuum (MacDonald, 2012)
- Have overarching audience research goals and questions, possibly shared with other organisations (see below)
- Staff responsible for evaluation need to explain its potential benefits and present it as an investment that can speed things up by answering key questions at the right time, rather than a cost that can slow things down (Wilkins, 2012)
- Funders can increase the impact of summative evaluation in a museum if they engage with grant-recipients' evaluation work: set high expectations of transparency and honesty; engage with the evaluation and ask questions of the report. Consider requiring plans for acting on the findings of each evaluation

ii) Have **mechanisms and structures** for raising awareness of evaluation, retaining knowledge from it and demonstrating its relevance to a range of staff. Possibilities include:

- Easily searchable in-house database or put evaluations (or key findings) on intranet

- Include briefings on evaluation findings at staff meetings; run lunch-time or all-day seminars on evaluation findings
- See the task as informing the rest of the organisation (rather than convincing them!) (Frost, 2012a)
- Work, for example through training, to make staff feel comfortable with evaluation and to recognise its benefits. Foster 'Better understanding and respect for audience research among other departments; eg from curatorial side "it's not real research".' (Anonymous contributor to final colloquium).
- Involve a range of staff in preparing the brief for an evaluation so their interests are taken into account
- Allow some staff to observe focus groups
- Involve staff in data collection and involve them in the evaluation in other ways (provide training if necessary)
- Pay attention to the readability of the report (film clips can help get messages across)
- Have a regular (at least) annual meeting, including any external evaluators who have undertaken work for the museum, to reflect on the implications of a series of recent evaluations

iii) Have an individual or department with **clear responsibility** for evaluation and visitor studies. The individual/department can be a 'champion' for evaluation and build up knowledge long-term and apply it.

iv) **Plan for change.** Build in time and money for remedial change, to act on the findings of an evaluation. However, 'Will funders allow a significant amount of funding to go into remedial work?' (Featherstone, 2012)

To increase impact on knowledge and learning across the sector

Potential ways of improving the impact of summative evaluation more broadly include:

i) **Sharing** the findings of evaluations. The ideal is 'A clear overarching mechanism to share findings with each other' (Sinclair, 2012). Components of this might include:

- Publishing articles of key findings of evaluations, or of a group of evaluations.
- More face-to-face sharing of findings at, for example, seminars
- 'Safe places' virtual or real-world, for confidential sharing of evaluation findings
- Museums could use the evaluation report as an opportunity to bring together practitioners to discuss the strengths and weaknesses of individual galleries

However, barriers to sharing include:

- Commercial confidentiality clauses are starting to appear in some evaluation contracts (King, 2012b)
- 'I resent the comment that "we should all publish". There is no time in a full-time museum job.' (Anonymous contributor to second colloquium)

- Designers have lots of knowledge and would like to share more but don't because it would take time and there may be issues of intellectual property with competitors taking ideas (Higgins, 2012a)
- Shrinking travel budgets make attendance at seminars and events harder
- As noted above, there is resistance to sharing failures or shortcomings

To overcome the barriers:

- Create a general sense of desirability of dissemination so that organisations build it in to the design of projects and include it in funding applications and in briefs to evaluators.
- Build time and resources into projects (or evaluation contracts) to allow museum staff (or evaluators) to publish findings or disseminate in other ways
- Funders, possibly working with sector bodies, could support, or undertake, dissemination
- Funders could help 'promote both positive and negative findings, so people could see the value of honest reporting' (anonymous participant in second colloquium).

ii) Consider the merits of **publishing** all summative evaluations. This could be encouraged (or mandated) by funders. The more easily available evaluations are, the more likely they are to be consulted, so encouraging synthesis. One example is work by students at the University of Washington to review evaluations posted on www.informalscience.org (Dewitt, 2012). Many of the reports selected for analysis in appendix 2 to the present report were found online. An on-line database would be most useful if it was easily searchable and reports were well categorised/tagged, ideally using consistent terminology.

However, the expectation of publication (and indeed other forms of non-confidential dissemination) could lead to less frank evaluation reports (Talbot, 2012). There is also a view that simple publication of evaluation reports is by itself not enough; what is needed is careful consideration of results. 'Not just better dissemination, but reflexive practice with appropriate resource to do it... Summative evaluation isn't useful more widely.... [for that] should focus on research questions, synthesis...' (Pegram, 2012b). Another participant in the second colloquium observed, 'We are all deluged with information and fight hard to ignore it... need a communication system that delivers the right information to organisations that need it at the right points, otherwise it will be ignored.' This is perhaps a counsel of perfection, but certainly a synthesis of multiple findings (see below) will often be better than raw reports.

iii) We have found wide support for the suggestion that there should be **syntheses** of findings (or 'distillations of learning') from a range of evaluation studies to identify common findings. This could be undertaken in a number of ways:

- An organisation or individual could be commissioned to synthesise findings and produce a 'meta evaluation'
- Museums (and possibly evaluators and funders) could collaborate with academic organisations to distil key learning from evaluations

- Funders could commission work to synthesise learning from evaluations (and other evidence) related to projects they have funded. This could be in the form of reports, perhaps commissioned from universities, or something closer to an audit of a range of projects. It has been suggested that funders could play a particular role in investigating longer-term impacts (Wellcome Trust, 2012)

Summative evaluation is rarely commissioned with the idea it will be useful to the wider sector. So, in the words of an anonymous second colloquium participant - and as we found in working on appendix 2 for this report - synthesis of existing summative evaluation reports is unlikely to produce a great deal, 'Synthesis and dissemination... would only work if evaluations were better commissioned in the first instance.' Key to this is specifying general research questions to address in addition to specific project objectives (see below). In addition museums and evaluators could:

- Consider standardising some aspects of data collection and methodology to aid comparability of evaluation studies

iv) There appears to be potential in identifying **general research questions** to give a wider context for summative evaluation. This could be approached in a variety of ways:

- Individual organisations could identify overarching research questions. The Science Museum has four: impact of audience co-creation and participation on the wider museum audience; audience engagement with objects; the role of new media in enhancing audience engagement; and impact of direct engagement of the public with scientists. These feed in to and influence the evaluations it commissions and help 'gather our findings from across different projects to contribute to a broader understanding with a variety of supporting evidence... that means that we might be in a better position to write up and share that learning once we've acquired it.' (Steiner, 2012) Tate is linking evaluation to its programme of 'research through practice' that runs through its learning work, 'constantly questioning the ways in which we work and the reasons why we do it' (Tate Learning, 2010)
- A group of organisations could share the same research questions
- A group of organisations could go further in taking a research approach and set up hypotheses to test in several different settings (Griggs, 2012b)
- Less formally, a group of organisations could agree to share findings in a particular area for a year or two, iteratively building up knowledge through their evaluations and culminating in a 'how to guide' like those previously produced by the Science Museum and Ben Gammon (for example Gammon, nd(a) and Science Museum, nd)
- Going further, there could be sector-wide research objectives, supported by major funders, or museums to apply in their own evaluation work (King, 2012b)

v) There is a general need to **raise awareness** of the potential benefits of evaluation, 'a better understanding of the benefits and importance of evaluation so that it is not just something that is 'done' but is embedded in the process and becomes as

important as design/marketing/interpretation etc' (Wilkins, 2012). Evaluators could work together, possibly through the Visitor Studies Group, to demonstrate to museums and others the potential benefits of evaluation and showcase examples of best practice including:

- good briefing
- exemplary evaluation studies
- examples where evaluation has had a significant impact on practice

One way to achieve this could be to work with federations and subject specialist networks so a wider range of staff see the value of evaluation and different curatorial groups feel evaluation is relevant to them. There is also a need to increase museum practitioners' skills in commissioning and understanding evaluation.

Appendix 2

Some principal findings of summative evaluation

Contents of Appendix 2

Introduction	46
1 Difficulties with comparing data	46
2 Gallery texts: panels and labels	49
3 Use of interactives	53
4 Use of other interpretive material	54
5 Visitor orientation	55
6 Learning	56
7 Visitor interests	58
8 Some dislikes, or fears	59
9 Time spent	60
10 Proportion of things viewed	63
11 Extending engagement	64
12 Conclusion	66

Summary

We attempted to identify a number of common findings and themes that arose across a range of summative evaluations. There are surprisingly few more general conclusions that can be drawn from the reports, which in many ways paint a rather disheartening picture of visitor (dis)engagement.

Labels Evaluation of permanent galleries paints a picture of typical visitors ‘grazing’ - being attracted to an object and then turning to the accompanying label. If interpretation is not available on labels adjacent to objects, many visitors do not seek it out elsewhere. Labels can successfully engage visitors more closely with objects. Simple texts appear to suit the majority of visitors. Labels aimed at specific audiences are not always recognised as such. There are numerous reports of visitor

comments that type size is too small to read easily. Some words or phrases commonly used in museum texts do not appear to be well understood by visitors.

Text panels There is a recurrent finding that only a small proportion of visitors to permanent galleries read text panels. However, visitors expect gallery panels to be provided. It does not help visitor orientation if introductory panels are placed only at one entrance to a gallery with multiple entrance points. Visitors to temporary exhibitions are far more likely to read text panels. However, a substantial minority do not read introductory panels, even in charged-for temporary exhibitions.

Orientation and understanding The arrangement and themes of a gallery often remain hidden to visitors. Interpretation can be missed by visitors if it is positioned or designed too discreetly. Evaluations do not provide extensive or convincing evidence of extensive or deep learning in galleries.

Time spent and attention paid Many people view a low proportion of the exhibits in permanent displays. It may be unrealistic to expect most visitors to engage with a large number of exhibits. Evaluations of large museums find surprisingly low dwell times in permanent galleries. It is quite common for the average length of time spent in a gallery to be only a few minutes. Visitors seem to stay longer in charged-for temporary exhibitions. Some, but by no means all, interpretive interventions can increase time spent and attention paid. Interacting with staff on the gallery floor appears to extend visitor engagement.

Interaction Visitors react positively to interactives, reporting that they enhance their experience. Visitors can be nervous about touching objects, including replicas, that have been made available for handling in galleries. Reading other visitors' comments is popular.

Visitor interests Evaluations have found a perhaps bewilderingly wide range of visitor interests. It is regularly found that visitors express a preference for interpretation that identifies connections to people, or with themselves.

In spite of the significant expenditure of time and money on evaluation there remain many areas that require further investigation if museums are to have a clearer, empirically-based, understanding of what makes a successful gallery.

Overview of some principal findings of summative evaluation

Introduction

Evaluation reports typically include a wide range of observations and findings concerning patterns of visitor behaviour, engagement with exhibits, the use of interpretative resources, participation and learning. Many include recommendations addressed to the particular museum under observation. In the following, we attempt to briefly summarise some of the key findings of a wide range of evaluation reports and to assess their more general implications for our understanding of visitor behaviour and the design of museums and galleries. However, as discussed in the main body of this report, rather little of the findings seem to have wider applicability, beyond the particular gallery or museum being studied.

The overwhelming sense after reading several dozen evaluation reports is one of disappointment that evaluations, as noted in an section 2c of this report, usually include so little about visitor behaviour and engagement at the 'exhibit face'. Indeed, there is a distinct sense in many evaluations of simply not knowing what the visitor is actually getting from the experience of visiting. A small number of studies do try to probe this more deeply but, with the exception of interactive science exhibits, evaluations tend to address the visitor experience at the level of the entire gallery or the entire visit, so do not help with understanding engagement with individual exhibits.

1 Difficulties with comparing data

In comparing the findings reported in different evaluation reports, it is important to be aware that, as noted in section 1 of the main body of this report, evaluations collect and report data according to varied protocols, so apparently comparable data might not in fact be comparable.

For example, an evaluation will typically have a rule for sampling, such as every fifth person to enter the gallery, but some protocols may then exclude certain visitors from the sample. For example, a study of a permanent gallery in a science centre excluded people who spent less than one minute in the gallery (Gammon & al, 2011, p. 51); an evaluation of a permanent museum gallery counted 17% of observed visitors as 'walkthroughs' but then appears to have excluded them from the data in rest of the report (Kuechler, 2009). These decisions are sensible, but will mean that the reports' dwell times are higher than they would if they included every selected visitor, as appears to be the case in some other evaluation reports. There also appear to be differing definitions of 'walkthroughs'. A rare example of an evaluation report giving a precise definition is given in the table.

Evaluation	Protocol	Notes
(Gammon & al, 2011)	Excludes visitors who spend under one minute	
(Kuechler, 2009)	Excludes 'walkthroughs'	Reports on the percentage of 'walkthroughs' in several galleries. Proportion ranges from 17% - 91%
(Mair, 2008)	Excludes walkthroughs and gives a clear definition (see right)	'straight from entry to exit without stopping to look at any element... many walkthroughs had cursory glances... but unless they stopped they were not recorded as having any engagement with the exhibition.' 99 of 154 observed visitors (64%) were classified as walkthroughs and so excluded from data analysis

The difficulty of precise definition can be seen in this account of the British Museum's overall approach, 'The issue of whether someone is counted a walkthrough or not depends on a judgement about their level of engagement. If someone walked through the display, and glanced at a display for a few seconds they would be classified as a walkthrough. There has to be some meaningful depth or duration of engagement. If someone came into the gallery, sat on the bench for five minutes, but didn't really look the displays, they'd be a walkthrough.' (Frost, 2012b)

There are also varying definitions of engaging with an object or item of text. Some observation protocols record a visitor engagement if the visitor merely glances at something; others require the visitor to physically stop in front of it (Harknett, 2012) (Mair, 2008). One report specified a stop of greater than 5 seconds (Fisher & Paynter, 2011). The decision to exclude stops of five seconds or less has a significant impact on reported findings – 57% of visitors looked at one key exhibit, but only 33% for greater than five seconds – in this case, the exclusion of short stops therefore roughly halves reported visitor engagement (Fisher & Paynter, 2011, p. 67). In an alternative definition 'A visitor engagement in an area/ display or an interactive was reported if a visitor had engaged with an area/display or interactive in a manner that was visible to the researcher.' (Fusion, 2011a, p. 7). Some evaluation reports set out a detailed categorisation of levels of engagement, but do not make it fully clear how, or whether, the categories are used in the analysis of data. One example is: L1 Glance; L2 <10 seconds; L3 10-40 secs; L4 >40 secs (Kuechler, 2009, p. 16(slide)). Another example is 'Orientation: Acknowledge the exhibit, but quickly moving on; Exploration: Briefly consider the exhibit; Discovery: Spend time looking at the exhibit in some depth; Immersion: Spending an extended period looking at exhibit in depth, studying it closely' (MHM, A Privileged Insight: An evaluation of Shah 'Abbas: The Remaking of Iran, 2009a, p. 27). One report that uses these categories (undertaken by a different group of

consultants) optimistically defines 'discovery' as anything greater than 10 seconds! The report calls the percentage of people who looked at each case (including glancing) the 'attracting power'. It also talks about the case's 'holding power', which is defined as the percentage of people attracted to the case who then stayed for over 10 seconds (Clinckemaillie, Lybarger, & Peterson, 2010, pp. 33-4).

A further factor to consider is that some evaluations rely strongly (or wholly) on visitors' own reports of their behaviour, although there is evidence that this is unreliable when comparisons are made with observation data. In particular, visitors tend to over-report time spent and attention paid. 'Similar to amount of Gallery seen, visitors in general tend to overestimate the amount time spent in a Gallery' (Fusion, 2011a, p. 9). For example, 32% of visitors to the British Museum Money Gallery reported spending over 15 minutes, observation showed in fact only 5% spend more than 15 minutes (Orna-Ornstein, c1998, p. 14). In another example, 35% of visitors were observed to not engage with a particular type of exhibit ('cases with objects') for more than five seconds, but when asked only 7% of visitors said they didn't see or use that type of exhibit (Fisher & Paynter, 2011, pp. 43, 46). (It may be that the discrepancy can be accounted for by people who accurately reported that they used the exhibit type but were observed to do so for five seconds or less; this would then show the dramatic ways reported results vary depending on the approach and definitions used.) At V&A Medieval and Renaissance Galleries 'interactive users report having seen a larger proportion of a gallery than non-users.' (Fusion, 2011a, p. 9) but observations found 'there is no correlation between the number of areas/displays and interactives and the time spent in a gallery' (Fusion, 2011a, p. 8). As with other aspects of engagement, visitors appear to over-report their reading of labels.

In a further problem with self-reporting, evaluators observe that visitors can be 'incredibly resistant to being negative.' (Griggs, 2012a) For example, people 'seemed reluctant to use the word "boring" and often used "indifferent" instead' (Bishop & al, 2006, p. 4). Visitors are in general perhaps easily pleased: when asked what they liked least in a survey of a museum, 'Over two-thirds of respondents indicated that there was nothing they liked least about the museum, indicating high overall levels of satisfaction' (Scotinform, 2008, p. 44).

Visitors may also interpret their behaviour differently to the researcher, which can lead to significant differences between self-reported and observed behaviour. For example in one gallery observation, 'emotions were much in evidence' but only 16% of those asked agreed that the gallery 'stirred my emotions'. The researcher suggests this may be because of differing understanding of what an emotional response might encompass: 'Why were visitors reluctant to recognise that their emotions were stirred...? Visitors were observed shrieking with panic,... bursting with pride,... profoundly moved... Emotions were much in evidence. This may well be a vocabulary issue... Would someone who was excited by the "driving backward" challenge describe his emotions as being stirred?' (Fisher & Paynter, 2011, pp. 28-9). Most of the results given below are based on observed rather than self-reported data, but there are still, of course, problems with comparing reports of observed behaviour.

2 Gallery texts: panels and labels

There is a recurrent finding that only a small proportion of visitors to permanent galleries read text panels. ‘The majority of visitors to free displays of the British Museum won’t read panels.’ (Francis, 2011, p. 158) ‘The attracting power of introductory and section panels was never higher than 10% and in many cases not a single visitor was observed to stop at them.’ (Francis, 2011, p. 154) This latter statement is slightly contradicted in another British Museum report that shows the ‘attracting power’ of permanent gallery introductory panels ranging from 2% to 44% (Kuechler, 2009, p. 10 (slide)). In the Assembling Bodies exhibition at Cambridge University Museum of Archaeology and Anthropology, all text panels have lowest rating for dwell time – one text panel was looked at by no one observed at all (Harknett, 2012). At the V&A permanent Jameel Gallery 25% of visitors were observed to read introductory panels (Fusion, 2008, p. 6). In formative evaluation at the Museum of London ‘Visitor views were mixed in terms of the importance of reading the introduction to aid understanding of the contents of the display case... In terms of their own understanding, most adults did not feel reading the introduction was necessary.’ They thought they would get the information they needed from individual object labels. ‘Nobody felt enjoyment of the objects and the display case in general would be compromised if an individual had not read the introduction’ (Creative Research, 2008b, p. 19).

People are more likely to read panels in temporary exhibitions. In BM charged-for temporary exhibitions, typically 60% or more of visitors read the introductory panel (Francis, 2011, p. 154); for example, 74% of visitors engaged with the Introductory text panel in the Shah ‘Abbas exhibition, for an average of 34 seconds (9% missed it because of queuing); 61% engaged with the introductory map, for a relatively lengthy 103 seconds (11% missed it due to crowding) (MHM, 2009a, pp. 29, 30).

However, a substantial minority of visitors to temporary exhibitions do not read text. In a charged-for highly interactive exhibition at the Natural History Museum around a third of visitors avoid it, although other visitors ‘asked for more and clearer instructions, particularly at the beginning’ (Pontin, 2007, pp. 17-18). Conversely, **in some permanent settings panels seem well used:** At the National Trust’s Dyrham Park, ‘in rooms with text panels all but one group observed read them’ (Boyd & Graham, 2010, p. 28).

In a temporary highly interactive exhibition at the Natural History Museum visitors read on average eight pieces of text in an average visit of over 30 minutes. Adults were more likely to read text than children. For example, most of the children in school parties read no text at all’ (Bishop, 2008, p. 6). Realistically ‘even those who stated a preference [for a particular panel, in a formative evaluation] may still not choose to read text panels under normal circumstances’ (James, 2000, p. 8). However, **visitors expect gallery panels to be provided:** almost 80% of visitors to the V&A ceramics galleries rated wall panels as ‘very important’ (Fusion, 2011b, p. 6). An answer to this dilemma might be to do very brief panels, essentially for orientation (Francis, 2011, p. 159).

At the BM ‘object captions are the most frequently read of all interpretive texts’ (Francis, 2011, p. 156). Over 60% of Glasgow Museum of Transport visitors agreed strongly (and over

30% agreed slightly) with the statement 'I like to read labels'. This varied slightly by age, with over 35s rating label reading higher than under 25s. Label reading was rated higher than any other form of interpretation by all age groups (Scotinform, 2008, pp. 33-34). When asked, most visitors 'considered it important... that every object has a corresponding caption or label' (Creative Research, 2008b, p. 11).

Evaluation paints a picture of visitors being attracted to an object and then turning to the accompanying label. In a front-end evaluation to test visitor reactions to objects being considered for display at the Natural History Museum, 'many visitors actively read the labels [which showed only factual information] before deciding how they felt' (Bishop & al, 2006). 'When asked if they would look at objects then study the captions for those they found interesting, look at captions then find the objects for those they found interesting or a combination of both, most [19 out of 27], said they would look at the objects and then try to find out more about them. Just 1 out of 27 said they studied the captions and for those that looked interesting tried to find the objects they referred to, while 7 out of 27 did a combination of both' (Creative Research, 2008b, p. 11).

Evaluations can give a misleading impression of a high level of use of gallery texts. At V&A British Galleries 91% agreed 'labels and panels helped me look more closely at the objects' (Creative Research, 2002, p. 21). At V&A Jameel Gallery, 'Approximately 2/3 of visitors were observed glancing or reading labels or at least part of them' (Fusion, 2008, p. 11). In the V&A British Galleries around half reported that they had read an introductory panel 'in the area they had just been in, or others like it in other displays' (McManus, 2003, p. 15). These relatively high levels of engagement may be because information is self-reported or because the object of study is panels or labels in general rather than any one individual panel or label.

The V&A sums up its learning about use of text, 'In reality, visitors don't always read gallery text as diligently as we might like. In exhibitions perhaps they do, but in permanent galleries they tend to stop and graze – reading a few labels and moving on, often ignoring the panels. So while every piece of text should link to the display, it should also be independent and make sense out of context.' (Trench, 2009, p. 14)

Labels can successfully engage visitors more closely with objects. In a BM Room 3 exhibition the majority of visitors went straight to the picture (before reading text) 'the directive text – actively encouraging visitors to look at elements of the painting – successfully increased engagement and learning and visitors liked being pointed towards these outcomes.' (MHM, 2008a, p. 36) In National Trust Houses, 'Where text pointed visitors towards things to see and do, they often followed it, looking closely at specific features or noticing aspects of displays.' (Boyd & Graham, 2010, p. 28) If visitors are to engage with an exhibit, it is of course necessary to attract their attention to it in the first place. Here, text can have a key role, 'It is crucial... to provide a narrative which can be easily apprehended by the visitor who is just scanning' in part this will come from 'pithy labels which raise the questions and issues.' (Fisher & Paynter, 2011, p. 56) However, merely changing the text in a case without addressing the overall layout had a 'negligible' impact (Clinckemallie, Lybarger, & Peterson, 2010, p. 56).

70% of visitors sampled during a formative evaluation preferred **simple text**. 'Chunking the text into smaller paragraphs was commented on favourably by several visitors.' (James, 2000, pp. 8, 9) The evaluation cites (but does not fully reference) Science Museum 1995 research that 'dividing a 150-word label into three 50-word labels increases readership from 11% to 35%.' (James, 2000, p. 16). In conjunction with interactive exhibits, 'If labels are short and clearly highlight key words and phrases, parents feel more supported and are more able to feed the information on to children'. Labels should perhaps not try to be too clever: 'Visitors interpret exhibits and labels very literally... Use of metaphors... must be done with caution.' (Davis, 2007, pp. 7, 9) In the Natural History Museum's Dino Jaws exhibition, the use of metaphor meant some children 'left the exhibition with the misconception that dinosaurs ate food such as sausages' (Pontin, 2007, p. 11). And **text should perhaps not be too subtle**, as visitors often 'do not understand clues that you may think are completely obvious' (Davis, 2007, p. 7).

The great majority of visitors to Kelvingrove - where text is relatively short and simple and aimed at five target groups of schools, under 5s, families, teenagers and non-experts - said the type and amount of text was 'just right'; around 10% said it was slightly too simple or far too simple. 'Most (but not all) of those saying too simple were in the AB social grade' (Carter, 2008, p. 45 (slide)). Even though some visitors were critical of the simplicity of some texts, 'People generally like the fact that the information is accessible to everyone... Objectively, general and specialist visitors to Kelvingrove, while wanting more detailed information from the museum for themselves, were very happy that it appealed to everyone and gave people an opportunity to learn more. This is a regular finding from specialist [visitors] in museums, as they are usually happy with a "general" level of information, with which they can confirm their own knowledge. Specialists are then keen for others to learn more about the subjects in which they are interested.' However, for themselves, 'experienced visitors want more information... At times the interpretation stops short of telling more experienced visitors what they want to know.' (MHM, 2009b, pp. 66, 64)

Formative evaluation suggests **children want plain, simple texts** (James, 2000, p. 14). 'Most of the children were overwhelmed by the [proposed] amount of text/information' in a mock-up of a panel for family-friendly displays in the Ashmolean Money Gallery (Nomikou, 2011, p. 173). Young people (9-14) visiting the Ancient House Museum, Thetford reported that they did not read labels (however, the evaluation report includes an example of the same young people seeking out a label on an object) (Playtrain, 2008, p. 10).

Other aspects of the style and content of text are rarely discussed. A review of a group of evaluation reports observed, 'Most of the evaluation reports failed to provide any great detail on the types of text used and how effective these were.' (Pontin, 2010, p. 8) **There are suggestions that the tone of text can make a difference**, 'If we would like our visitors to respond positively to our displays, we have to show our own love for the collections. Recent surveys show that V&A visitors sometimes feel that the authors of our text are remote and stand-offish.' (Trench, 2009, p. 35) **Labels aimed at specific audiences are not always recognised as such**. In one case of labels designed differently for children 'adults did not always recognise the caption as one that was aimed at children (adults in family group were more likely to notice this) although they recognised it was different.' It is therefore

important to 'Ensure any captions aimed at children and families are instantly recognisable to children.' (Creative Research, 2008b, pp. 9, 41) Fuller interpretation might be advised for unfamiliar exhibits, such as art in a science museum (where in fact there may be less interpretation than in the rest of the museum) (Dawson, 2006, p. 11).

Some words or phrases commonly used in museum texts do not appear to be well understood by visitors. Examples noted in evaluations include: 'iconic' (Bishop & al, 2006), 'arachnid' (Creative Research, 2008a, p. 18), 'taxonomic', 'type specimen' (Creative Research, 2005, p. 55) and 'self-guided tour' (Creative Research, 2008a, p. 11). Visitors displayed 'some uncertainty' with terms such as 'biodiversity', 'digitisation' and 'natural science' (Creative Research, 2005, p. 55). Occasionally, formative evaluations consider the comprehensibility of language to children (Ironsides, 2009, pp. 15, 21).

Some evaluations comment on the format or positioning of text. In one British Museum exhibition there was praise from visitors for 'the tactic of repeating labels on different sides of exhibits allowing for a maximum number of visitors to view an object at any one time, although some labels were felt to be positioned too low down' (MHM, 2009a, p. 42). Visitors report that when labels are grouped together, at the bottom of a case on a 'caption rail' they are happy with photographs of objects on labels to link them to the objects, as long as the photographs do not make the object hard to see. Photographs on labels 'did not discourage people from looking at objects.' For ease of reference, captions and labels should align with the objects they refer to; when labels were not aligned, 'a few [visitors] seemed particularly put off by this and sometimes continued to mention it' (Creative Research, 2008b, pp. 4, 14). Visitors to the V&A British Galleries 'encountered no problems in using object numbers in order to help identify labels' (McManus, 2003, p. 16).

If interpretation is not available on labels adjacent to objects, many visitors do not seek it out elsewhere. Evaluation of one BM (free to enter) exhibition found that even once engaged with an exhibit, 30-50% of visitors did not turn to relevant nearby text panels when seeking information about minimally labelled objects. In a BM Room 3 exhibition, interpretation located behind the object was missed by a lot of visitors (MHM, 2008a). Fewer than 5% used V&A Ceramics Gallery 'search the collection' computers (Fusion, 2011). In a BM free exhibition 'around half of the visitors were attracted to an object; looked for the nearest piece of information, that is, the label; found that it did not answer their immediate questions, and left... object information should always be placed next to the object.' (Francis, 2011, p. 158) Text panels will be ignored if they are hard to read. In the final Reading Room' of Tate Britain's 2008 Turner Prize Exhibition only 15% read the wall panels, which were 'consistently unpopular across all age groups... This might be explained by the location of the panels behind the chairs used for watching artist videos, or the colour of the text (light grey on white). In future, the wall panel texts would benefit from being placed in a more prominent position, and would need to be presented in darker type to ensure that visitors can read them easily and at a distance.' (Honeybourne, 2009, p. 4)

It does not help visitor orientation if introductory panels are placed only at one entrance to a gallery with multiple entrance points. Visitors waste time and energy looking for the introductory panel, or fail to get the 'proposition' of the gallery (MHM, 2009b, pp. 53-4, 57).

Introductory text on a caption rail at the bottom of a case was not always recognised as such by visitors, who did not expect it to be located there; if the case was busy with visitors, views of the introductory text could be blocked and it might not even be noticed (Creative Research, 2008b, pp. 5, 18).

There are numerous reports of visitor comments that type size is too small to read easily.

One report includes a suggestion from members of an access advisory panel that ‘black text on white background is difficult to read for some people with dyslexia, as the contrast is too high’ (Mair, 2008, p. 4).

A number of questions remain largely unaddressed in many evaluation studies concerning the use of texts that would seem critical to understanding their use and significance. For instance, when people visit with others (a large proportion of visitors in many cases), it is not unusual to find particular individuals reading the label and summarising its content for others, so a relatively low ‘reading rate’ may not necessarily correspond to low impact of the label. Secondly, the structure and form of the label (or content provided by a kiosk, PDA, etc) has an important impact on how the information is used and how visitors engage the exhibit(s) both alone and when with others, and yet this remains largely unexplored.

3 Use of interactives

Probably the best evaluated area of museum activity at the exhibit-face-level is science-based interactive exhibits (for example, Gammon, 2011, Humphrey, 2005, Davis, 2007). However, here, we largely focus on evaluation findings about interactive exhibits used to enhance interpretation of object-based displays. In the V&A Medieval and Renaissance Galleries ‘40% of visitors report using a gallery interactive during their visit’; observation shows this varies from 26%-76% depending on the gallery (Fusion, 2011a, pp. 9, 8). At V&A British Galleries, 57% used at least one interactive area, including videos (Creative Research, 2002, pp. 19-20). At the V&A Jameel Gallery about half were observed using interactives; in an unusual example of visitors under-reporting engagement, about a third said they used them when asked in interview (Fusion, 2008, p. 16). **Having more interactives doesn’t necessarily mean people use more of them.** ‘The number of interactives [in a V&A Medieval and Renaissance gallery] does not correlate with higher interactive engagement.’ (Fusion, 2011a, p. 9)

Visitors react positively to interactives, reporting that they enhance their experience. At V&A British Galleries ‘Among those using at least one interactive exhibit, 93% felt they enhanced their appreciation of the objects and 89% felt they helped improve their knowledge of the subject.’ (Creative Research, 2002, p. 14) At BM Hands On desks, 96% said it increased the quality of their visit. Some people didn’t mind replicas; others did; some were nervous of touching (MHM, 2008b, p. 37). Evaluation at the Churchill Museum found ‘considerable evidence that multimedia encouraged visitors to learn in different ways. They looked in more depth than they would otherwise have done, to actively search information and to learn alongside other members of their group.’ However, ‘some find too much noise and IT confusing.’ (Pontin, 2010, pp. 9, 10)

As an example of evaluation findings about visitor engagement with wholly interactive galleries, in *All About Us at @Bristol*, over a third of visitors had 'complete interaction' at three or more exhibits, which suggests around 2/3 of visitors had 'complete interaction' at fewer than three (out of over 50). Typically half of visitor interactions were 'incomplete' (Gammon & al, 2011, pp. 58-9). School groups managed to successfully operate only around half of the exhibits they attempted to use, even on accompanied visits with a facilitating adult (Gammon & al, 2011, p. 60). At the Science Museum's *Fuelling the Future* 'About half of visitors do not start their interaction with exhibits at the beginning of the exhibit's sequence. This is due to visitors "taking over" these exhibits when the previous visitor leaves before the exhibit has returned to the attractor screen.' This means they miss the introduction and in some cases the interaction does not work properly (Hayward, N & Burch, A, 2005, p. 5).

Comparing and contrasting the observations and findings from evaluations that assess 'interactives' poses particular difficulties. There is a substantial range of different types of interactive found within museums and galleries that range from stand-alone computer based exhibits, through to simple mechanical or material assemblies in which for example a visitor might be required construct a bridge or chair. They include stand-alone exhibits as well as interpretative resources such as video that illuminate the operation of particular object on display. Moreover, the 'content' associated with particular types of 'interactive' will vary significantly, and the content rather than the forms of 'interaction' the device or display provides, will have a profound impact on the experience of visitors. To make matters more difficult still, the location of the interactive within a gallery of objects, or other interactives, will bear not only upon its use but also on its effect on the visitor's engagement within other exhibits in the same gallery. Despite the wide-ranging corpus of evaluation reports that address interactives many important questions remain unaddressed and there is a necessity to identify some key distinctions to enable comparative analysis to be undertaken.

4 Use of other interpretive material

In the British Museum *Shah 'Abbas charged-for* exhibition 53% took the exhibition leaflet 'however, they did not necessarily use it during their visit. Just under a third used the audio guide. Some visitors chose not to use it because they 'were under the impression that the audio guide was uninterruptible and would dictate their viewing order' (MHM, 2009c, pp. 43-44).

As noted above, fewer than 5% used V&A Ceramics Gallery 'search the collection' computers (Fusion, 2011a). In a free-to-enter exhibition at Glasgow Museum of Transport, 'The most surprising result from the tracking was that not many visitors actually made use of the AV screens.' Excluding people who glanced at AV screens, only 22% of visitors watched an AV, for an average of 38 seconds. Very few people observed watched more than one AV and no one watched an entire AV. 'Given the number of AV screens in the exhibition and the level of interpretation they provide, the low level of use is quite disappointing. But on the other

hand, it could be evidence that visitors find the objects engaging enough in their own right.’ (Mair, 2008, p. 16)

In rooms with enhanced interpretation at National Trust Houses ‘curious minds’ visitors were most enthusiastic about: room sheets and guide books, talking with room guides and ‘close access’ to replica documents and photos, ‘although they were not comfortable with handling them.’ Visitors were ‘keen to share their own knowledge’ with room guides (Boyd & Graham, 2010, pp. 4, 29).

A review of evaluations at the Imperial War Museum observed that some people, especially families, liked to have the opportunity to **comment and to read other people’s comments** (Pontin, 2010, p. 9). Reading other visitors’ comment cards was by far the most popular form of interpretation in the final ‘Reading Room’ of Tate Britain’s 2008 Turner Prize exhibition. 50% read comment cards, 20% watched videos, 15% wrote their own comment cards and 10% read books. Under 35s were more likely to write comments than older visitors. ‘What visitors gained from reading the comments of others has yet to be evaluated.’ Positively, ‘the reading of comment cards led to many discussions... between visitors about the exhibition... with the comments acting as a springboard for discussion’ (Honeybourne, 2009, p. 4).

The fact that over three times as many people read visitors’ comment cards as read the ‘official’ wall panels perhaps suggests there are possibilities for incorporating gallery-written interpretation amongst comment cards!

The review of evaluation at the Imperial War Museum suggests visitors take pleasure in being able to take something away, to remember and reflect on the visit (Pontin, 2010, p. 5). For an exhibition at the Natural History Museum, two thirds of visitors knew they could later use their exhibition tickets to log on to the exhibition website and 58% rated the idea as 4 or 5 (out of 5); however, just 14% of visitors in fact used the ticket on the website (Pontin, 2007, p. 15).

At the V&A British Galleries **many people missed interpretive aids** such as gallery books and fibre-optic torches ‘since so many people did not even notice these interpretive aids it is suspected that discreet design values and lack of inviting, obvious signage is a problem’ (McManus, 2003, p. 17). In the V&A Jameel Gallery ‘The interpretive devices should be made more prominent so that people see them and interact with them. Perhaps a notice needs to be added next to them to draw people’s attention to them’ (Fusion, 2008, p. 27).

5 Visitor orientation

The arrangement and themes of a gallery remain hidden to many visitors. This is unsurprising in view of evidence presented above that many visitors ignore introductory panels and AVs, or use entrances that do not have an introductory panel. At Kelvingrove, each gallery’s ‘proposition’ could be better explained as visitors don’t always ‘get it’. **Visitors who are unsure of a room’s proposition often find it difficult to appreciate and understand what they are looking at in the room,** and why certain objects are being displayed in that

particular gallery. Indeed, many Kelvingrove visitors did not realise that the organising principle of the entire museum is that objects are grouped to present 'stories' - until it was pointed out to them by the evaluators. The evaluation report suggests setting out each gallery's stories on an introductory panel in the gallery.

After visiting the British Galleries at the V&A, 33% of visitors were unaware that there is an intended start point (Creative Research, 2002, p. 42). Only around half 'claimed to have become aware of some themes during their visit' and around a third 'did not recognise the given themes at all' (McManus, 2003, p. 13). Visitors to the V&A Jameel Gallery did not seem to understand the themes around which the gallery was organised. Half said they could not identify specific themes (Fusion, 2008). 'None of the visitors interviewed seemed to understand the way the gallery was organised.' (Fusion, 2008, p. 16) In a British Museum charged-for exhibition 'few visitors understood that the exhibition was divided into four sections' (MHM, 2009a, p. 49). In a Science Museum gallery 'The four zones and their differing themes were not obvious to visitors. [They] roamed happily across the gallery trying out exhibits which appealed to them.' However, 'even while they were enjoying themselves people were feeling a bit lost. Many couldn't quite see what it was adding up to.' (Fisher & Paynter, 2011, pp. 3, 33) Asked about plans for a new Egyptian display, people wanted each section or gallery to be clearly titled (Boyd & James, 2009, p. 36). On arrival at Kelvingrove very few visitors looked at the floor plan or got a gallery guide (MHM, 2009b, pp. 43, 57, 60, 68). Over half of visitors to the Pitt Rivers Museum didn't reach the two upper galleries and so, of course, failed to have any engagement at all with the exhibits there. (James, 2002)

6 Learning

Assessing visitors' learning presents severe methodological difficulties, 'A lot of the things museums do well in terms of working with audiences are fiendishly difficult to measure.' (King, 2012a) And there is of course the question of what constitutes learning. All teachers taking school children to an exhibition at the Natural History Museum saw the visit 'as an affective rather than cognitive learning experience' (Bishop, 2008, p. 8).

Some evaluations ask visitors to report whether they have learnt anything, but this information is next-to-useless as it is likely that out of politeness or embarrassment, when asked most visitors will tend to assert that they have learned something. (Nevertheless, there are examples in which as many as 50% of those asked report that they have not understood an exhibition – for example (Dawson, 2006, p. 9).)

An evaluation of a Science Museum gallery about energy reports 'strong evidence for learning' but in fact each of the four main gallery messages were identified by only 15% to 45% of visitors. Perhaps more encouragingly 'more than half of families felt inspired to learn more about energy having visited' (Hayward & Burch, 2005, pp. 9-11). In a Natural History Museum exhibition 'there is evidence that learning did occur but at a rather basic level... It is difficult from the evaluation... to define why visitors did not learn more.' (Pontin, 2007, p. 2)

Some evaluations use personal meaning maps (PMMs, also sometimes known as concept maps) to assess learning. At Pitt Rivers Body Arts the evaluation sees evidence of learning as after visiting the gallery ‘several’ visitors made corrections or added words to the PMMs they had prepared before visiting the gallery, but less than half improved ‘mastery’ (James, 2002, pp. 8, 22).

An alternative approach to investigating visitor learning may be to analyse comment cards that ask a specific question linked to the museum’s learning objectives (Mair, 2008, p. 33). Investigation of comment cards in Tate Britain’s 2008 Turner Prize exhibition found almost 50% of comments by observed visitors were negative and under 30% positive (the rest were neutral). Overall 45% of comments were ‘irrelevant’, not discussing the exhibition, Tate, or even art in general (Honeybourne, 2009, p. 4).

Before the V&A Jameel Gallery opened people held very stereotypical views of Islamic Art ‘limited to widely held but not necessarily correct views’ (Audience Focus, 2004, p. 14). At the Jameel Gallery as realised all visitors but one felt the exhibits and interpretation somehow improved their knowledge of Islamic Art (Fusion, 2008, p. 6). However on closer investigation ‘respondents did not seem to be able to elaborate extensively on concepts and did not demonstrate a very deep understanding’ (Fusion, 2008, p. 5). Comparing front-end and summative Jameel Gallery evaluations shows, in an unusually critical piece of published evaluation, ‘visitors reported similar things in relation to Islamic Art whether they had visited the exhibition or not... The gallery does not provide sufficient help with furthering visitors’ pre-existing knowledge and enhancing their understanding of more challenging and complex issues of Islamic art... visitors did not seem to be able to assimilate the information presented and therefore reach a deeper and higher level of knowledge and understanding... The sample for this study was consisted of well experienced museum visitors, several of whom were interested in art or had art-related qualifications... it was all visitors and not solely the self-confessed “novices in Islamic art”, who were not capable to clearly demonstrate a deep understanding of Islamic art and identify complex concepts.’ (Fusion, 2008, pp. 13-16)

In sharp contrast, the published summary evaluation of the V&A Medieval and Renaissance galleries states (without supporting evidence), ‘Visitors gained a comprehensive understanding of the period.’ (Fusion, 2011a, p. 7) The different conclusions of these evaluations is remarkable. Can it really be that the Medieval and Renaissance galleries are so much more effective at stimulating visitor learning. If so, why? In fact, it is hard to believe that visitors did gain a ‘comprehensive understanding’ when, just as one example, visitors spent an average of only 3 minutes 49 seconds in Gallery 64, Renaissance Art and Ideas 1400–1550 (Fusion, 2011a, p. 8).

In rare examples of an evaluation reporting **negative learning impacts**, at All About Us @Bristol when it was unclear to parents how to use an interactive, children would take home the negative message that something is hard to do (Gammon & al, 2011, p. 87). The Natural History Museum’s Dino Jaws exhibition inadvertently left some visitors with the erroneous impression that Dinosaurs ate sausages! (Pontin, 2007, p. 11)

In Who am I? at the Science Museum 'Learning feels fragmented... Visitors are left with indelible fragments and insights. They are primed to convert these into thinking about the wider world. However, this seems not to happen within the gallery itself.' (Fisher & Paynter, 2011, p. 2) 'Only a minority seemed able to identify and latch onto coherent scientific ideas' although 'most people discovered something close and personal about themselves and were wowed... They were excited and absorbed in themselves; it seems this was not the moment for stepping back and considering the whole. This may or may not come later.' Intriguingly, the evaluator's post-visit focus groups proved a good opportunity for visitors to reflect and make sense of what they had experienced, but the reflection did not happen earlier and might not have happened without the focus group' (Fisher & Paynter, 2011, pp. 33, 35-6).

Some evaluations consider visitors' prior knowledge of the subject covered by the museum, finding, for example, 'most visitors had only a basic knowledge' (Pontin, 2010, p. 7). (Incidentally, visitors with a high level of previous knowledge may be likely to answer 'no' to questions about whether their knowledge or understanding increased (Boyd, 2011).)

There remains much debate as to how it is possible to assess and measure learning and different methods of data collection and analysis produce highly variable results. In general, **we suspect there is a highly optimistic orientation to reporting learning that arises within museums and galleries**, and while it is rewarding to secure positive results, we are not at all confident that they would stand up to scrutiny or be treated as reliable by external organisations. It is also worthwhile considering how we might establish analysis of the longer term impact of museum visits and explore further ways in which continuing engagement with a particular museum or gallery might be facilitated.

7 Visitor interests

It is regularly found that visitors express a preference for interpretation that identifies connections to people, or with themselves. For @Bristol Science Centre, people are interested in information directly relevant to themselves (Gammon & al, 2011, p. 31); at the Science Museum Energy Gallery 'It appears that presenting things which were personally relevant provided an effective route into increasing families' interest in energy' (Hayward & Burch, 2005, p. 10). At the Pitt Rivers Museum, visitors want contextual information 'how things were worn or used', want to identify with people behind the objects, who made and used them (James, 2000, p. 15). 'Research at... displays and exhibitions at the [British] Museum has shown that visitors are always keen to hear the personal stories behind an object so that they can relate to their own lives.' (MHM, 2008a, p. 27) V&A text guidelines report 'We know from the Getty and other research [no sources are given] that people connect with people. This presents a problem in museums, where objects have been divorced from people.' (Trench, 2009, p. 24) At BM Hands On desks 'Visitors seemed to have the strongest rapport with items that connected them to people from the past.' People also liked things that were simple, very old, with a puzzle, or linked to personal interests (MHM, 2008b, p. 38). At V&A Medieval and Renaissance galleries 'many visitors enjoyed the everyday objects.' (Fusion, 2011a, p. 6)

Some visitors were disappointed that the British Museum Shah 'Abbas exhibition did not include social history and had an apolitical tone. A significant percentage want to better understand modern-day Iran (MHM, 2009a, pp. 6-7). At the Imperial War Museum 'visitors wish to see different points of view, be challenged and consider controversial subject matter... [They] generally appreciated links to modern society... However,... some visitors were doubtful about including current war.' (Pontin, 2010, pp. 11, 13, 14) 'The use of stories was strongly supported by all visitors', especially personal stories.

Interests vary between visitors. Overall, IWM evaluations have found a perhaps bewildering range of visitor interests, including basic facts, broader context, a local view, a clear timeline, narrative themes, the cost of war in its broadest sense, human experience, personal connections and to be challenged with differing points of view (Pontin, 2010, pp. 7, 8, 11, 12). Natural History Museum visitors 'have a wide range of interests' (Bishop & al, 2006) and at Pitt Rivers Museum 'Many visitors' views are directly contradictory. This makes it difficult to see the way forward. A decision needs to be made regarding where the museum's priorities lie.' (James, 2000, p. 17)

Unusually, an evaluation at the Ashmolean included a focus group with students and members of an archaeological society. They had rather different preferences for interpretation and display: detailed excavation numbers, bibliographic information, dense study displays. They also expressed a preference for some context and narrative (Boyd & James, 2009, pp. 9-21).

Cambridge University Museum of Archaeology and Anthropology has found pictures of people, life-size costumes, noisy or moving objects score well for visitor attention; there is also a 'halo' effect with objects/exhibits near to popular ones (Harknett, 2012). Evaluation of the Science Museum's Who Am I? suggested exhibits were more popular if they were close to an entry point, had something to catch the eye or ear, covered a subject 'which is intelligible to me'. Attractive objects are likely to be living/organic, large and dramatic. Many objects were perceived as banal. 'This meant visitors did not linger to grasp the underlying story... Single objects, uncluttered and touchable, 'were successful in getting people to pause and think in a relaxed informal way.' (Fisher & Paynter, 2011, pp. 51, 56)

8 Some dislikes, or fears

Visitors to Glasgow Museum of Transport are reported to feel self-conscious when looking at art as they feel their behaviour is being judged (Urquhart, 2001, p. 7). **Several evaluations report parents' fears that their children will damage things or behave inappropriately and visitors can be nervous about touching objects** or replicas that have been made available for handling. In some National Trust Houses 'removing ropes where visitors are still not allowed to physically enter spaces... has caused stress to visitors and staff alike' (Boyd & Graham, 2010, p. 5).

9 Time spent

Evaluations of large museums find surprisingly low dwell times in permanent galleries. It is quite common for the average length of time spent in a gallery to be only a few minutes. For example, at the V&A the average time spent in each ceramics gallery is 1-8 minutes (Fusion, 2011a); in the V&A Medieval and Renaissance Galleries average dwell time ranges from 2 minutes 40 seconds to just over 4 minutes (Fusion, 2011a, p. 8). In the V&A Jameel Gallery it is 4 minutes (Fusion, 2008). Similarly, at the British Museum, varies from just over one minute to around seven minutes (Kuechler, 2009, p. 18(slide)). Average dwell time for older permanent galleries is about 3 minutes (Francis, 2011, p. 159), For example, visitors spent an average of just over 3.5 minutes in the Money Gallery (Orna-Ornstein, c1998, p. 14). To contextualise this, the BM has over 90 permanent galleries and the average length of a visit to the BM is 2 hours 14 minutes. Not only does this include potentially 90+ free-to-enter galleries and charged-for permanent exhibitions, but also shops, cafés and toilets (Francis, 2011, p. 155).

Interactive galleries may get higher visit lengths. The permanent, charged-for interactive All About Us gallery at the @Bristol science centre has a mean dwell time of 17 minutes 41 seconds; median 13 minutes. About 20% of visitors spend 30-50 minutes (no one longer than that). According to the evaluator, 'These dwell times compare favourably with exhibitions of similar content and size at other UK institutions.' He also points out 'it was only practical to observe visitors for the duration of one visit to All about us, when in fact during an entire day they may revisit particular areas of the floor.' (Gammon, 2011, p. 38, 55)

Evaluations may give misleading results. The evaluation of the permanent Pitt Rivers Museum display of body arts found 'dwell-time varied between five and 45 minutes, with half the sample staying for 15-25 minutes.' (James, 2002, p. 6). That sounds high - the lowest reported time is greater than the average for BM permanent galleries. In fact, the sample were all completing pre- and post-visit concept maps and knew they were being observed. Visitors had been asked 'Could you spare some time to talk to me about your views on Body Arts? It should take about 5 minutes now. And then I'd like to talk to you again for about 5 minutes after you go through the display... You can look for me or my colleague as you leave the display.' (James, 2002, p. 14) The evaluation report comments 'the presence of the student data collector is likely to have influenced the visit to some extent, in spite of efforts to the contrary... In both studies the person approached often broke away from the rest of their group thus changing their visitor experience. Although this was not the intention, some visitors may have seen the Body Arts evaluation as a test of their knowledge and how much of the display content they could remember, thus influencing the way in which they approached the displays.' (p. 5)

Visit lengths to individual galleries

Place	Average time spent (rounded to whole minutes)	Notes	Source
Ceramics galleries, V&A	1-8 minutes (depending on gallery)		(Fusion, 2011a)
Medieval and renaissance galleries, V&A	3 - 4 minutes (depending on gallery)		(Fusion, 2011a, p. 8)
BM permanent galleries	1-7 minutes (depending on gallery)		(Kuechler, 2009, p. 18(slide))
Who am I? Science Museum	21 minutes	About 50% spent under 17 minutes; 26% spent over 28 minutes. Families spent significantly longer than independent adults	(Fisher & Paynter, 2011, pp. 2, 65)
Fuelling the Future, Science Museum	17 minutes	Maximum observed: 103 minutes	(Hayward, & Burch, 2005)
Digitopolis, Science Museum	11 minutes	Maximum observed: 46 minutes	(Hayward, & Burch, 2005, p. 4)
Body Arts, Pitt Rivers Museum	50% spent 15-25 minutes	May be artificially high; see text	(James, 2002, p. 6)
All About Us, @Bristol	median 13 minutes, mean 18 minutes;		(Gammon, 2011, p. 38)
British Galleries, V&A	median 60 minutes, mean 82 minutes (figures for people who had visited both floors at time of interview). Averages at under 4-5 minutes per gallery (allowing for time spent in interactive areas)	A sequence of 15 galleries and 8 interpretation areas. 'Respondents were asked to estimate how much time they spent in the British Galleries up to the point of the interview', so may under- or over- report	(Creative Research, 2002, pp. 70-71)

Visit lengths to entire museums

Where	Average visit length	Notes	Source
British Museum	134 minutes		(Francis, 2011, p. 155)
Glasgow Museum of Transport	94 minutes	Self-reported? Note: almost identical result found in 2003 and 2008	(Scotinform, 2003, p. 11); (Scotinform, 2008, p. 13)
Kelvingrove	100 minutes	About 50% of visitors spend 60-120 mins; about 20% spend over 120 mins	(MHM, 2009b, pp. 46-47)

Visitors seem to stay longer in charged-for temporary exhibitions. The BM's permanent Mexico Gallery has 160 objects and 12 panels and an average dwell time of 3 minutes 25 seconds; similarly, the British Museum's temporary, charged-for Moctezuma exhibition had 130 objects and 26 panels, although covering a larger floor area. In great contrast, the average dwell time in the exhibition was 79 minutes. This has been analysed in terms of differing motivations between visitors to the exhibition and to the BM as a whole (Francis, 2011, pp. 154-5). At the temporary but free-to-enter Assembling Bodies exhibition at Cambridge University Museum of Archaeology and Anthropology the average dwell time was 14.5 minutes (Harknett, 2012). Many evaluations cite research by Beverly Serrell that suggests that the average dwell time in any exhibition is around 13 minutes. More recently Serrell has given a figure of an average of 20 minutes for an exhibition visit (Serrell, nd c2010).

Visit lengths to temporary exhibitions

Where	Average visit length	Notes	Source
Ice Station Antarctica, Natural History Museum	36 minutes	Observed visits varied from 11-65 minutes	(Bishop, 2008)
Dino Jaws, Natural History Museum	27 minutes	'quite a long time for younger visitors'	(Pontin, 2007, p. 9)
Shah 'Abbas, British	Around 80% spent	Check whether self-	(MHM, 2009a, p. 26)

Museum	60-120 minutes	reported	
Lives in Motion, Glasgow Museum of Transport	1 minute 11 seconds	Exhibition free to enter and on a through route. Includes all who stopped to look at least one element of exhibition	(Mair, 2008, p. 14)
Reading room in Turner Prize exhibition, Tate Britain	60% spent under 2 minutes	NB this is for the final reading room only, not the entire exhibition. Figure appears to include walkthroughs	(Honeybourne, 2009, p. 3)

A number of questions and problems arise when attempting to summarise some of the key observations and findings of these various reports. First and foremost as suggested earlier, since the data is collected using different techniques and subject to different forms of analysis, it is not always clear that one is able to compare like with like. Secondly, a range of reasons may account for the time spent in particular galleries and with particular exhibits, including for example the overall number of visitors in the space at any one time, their engagement with particular exhibits, whether the particular gallery is the principal reason for visiting the museum (or their passing through to other galleries), and in consequence it is hard to draw firm implications. Thirdly, to a large extent, the forms and qualities of the behaviours that arise at exhibits remain largely unaddressed and there are no firm grounds to conclude that length of time is an indicator of engagement or participation.

10 Proportion of things viewed

Evaluations reveal a general behaviour pattern of browsing, with most (but not all) visitors viewing individual exhibits, rather than systematically following a sequence or theme. With that in mind, it is not surprising that evaluations also find that **many people view a low proportion of the exhibits in permanent displays.**

As an example, at the BM the average number of objects viewed per permanent gallery is just four (Francis, 2011). In the BM's Roman Empire Gallery the single most noticed object was looked at by just 14% of visitors; the 'most discussed' object in the same gallery was talked about by just 5% of visitors (Carlyle, Glover, & Klebanov, 2008, p. 38).

The recent evaluation of the V&A Medieval and Renaissance gallery relied on visitors self-reporting their engagement, so figures are likely to be unreliable. 58% of visitors report viewing 'half'; 17% viewed 'less than 25%' (Fusion, 2011, p. 6).

In All About Us, there are 7 sections. Half of visitors visit three or fewer sections; some sections are visited by fewer than 40% of visitors and some individual interactives were used by under 10% of visitors (p. 61). Of 50+ interactive exhibits, visitors used on average only 7 and, as noted below, they successfully completed only around half of even that small number (Gammon & al, 2011, pp. 58-9).

11 Extending engagement

Some, but by no means all, interpretive interventions can increase time spent and attention paid. Some modifications or interventions have been made to attempt to increase visitor engagement. The Getty Museum offers family-group visitors ‘art detective’ cards, focusing on a selection of artworks. An evaluation found families using the cards spent longer at related artworks than other families did at any artwork and that 70% of families using the cards spent over 2 minutes at a related artwork, compared to 43% of families without the cards (Jackson, 2006).

The British Museum introduced ‘gateway object labels’ in some galleries. This appears to increase dwell time. For example, Japan gallery visitors average 11 minutes and 9 objects – better than the BM norm of around 3.5 minutes and 4 objects (Francis, 2011, p. 159). However, there may be other factors: the Japan Gallery is at an edge of the British Museum site and is not on a through route, so as with charged-for exhibitions, visitors may have different motivations to general BM visitors.

Does interaction increase time spent in a gallery that mainly displays objects? At the V&A Medieval and Renaissance Galleries ‘there is no correlation between the number of areas/displays and interactives and the time spent in a gallery... The gallery with the highest interactive engagement rate had the lowest dwell time.’ (Fusion, 2011a, p. 8) (Note, however, that the gallery with the second highest interactive engagement rate had high dwell time.)

80% of users of ‘Hands On’ desks within British Museum permanent galleries spend under 5 minutes there (MHM, 2008b, p. 17). However, ‘it is difficult to say with any certainty whether looking at items at Hands On desks encourages visitors to look at more exhibits.’ 61% of visitors to the desks remain in the gallery afterwards (MHM, 2008b, p. 39).

Simple things, such as including seating can extend the amount of time people spend in an area (MHM, 2009a, p. 34), although they may not be engaging with exhibits for much or any of that time. At the Assembling Bodies exhibition at Cambridge University Museum of Archaeology it was investigated why visitor observation showed some exhibits to attract more attention than others. In one example of a Mongolian Household Chest it was found that the exhibit included a mirror which people were using to check their hair. In another example, there was somewhere to sit in front of the exhibit (Harknett, 2012).

Interacting with staff on the gallery floor appears to extend visitor engagement. In Ice Station Antarctica, an exhibition at the Natural History Museum: ‘Face to face interactions

with a facilitator did have a positive impact on visitors' experience and learning.' When asked, 'the vast majority of visitors felt that the facilitator made a positive impact on their visit both in terms of both increasing their knowledge and their enjoyment of the exhibition... In terms of increases in learning, a minority of visitors (16%) mentioned that they would now look more closely at specimens and objects in the Museum. A further 5% felt that they would remember more... Visitors who had a face to face interaction were more likely to stay longer in Ice station Antarctica than those who did not.' (Bishop, 2008, pp. 21-25)

In National Trust Houses that have been part of the 'Bringing Properties to Life' (BPTL) initiative visitors 'on average spent more time, engaged more closely with things and engaged in more "interpretive talk" in BPTL rooms than in more traditional rooms... the active participation of room guides was key.' In most rooms at least half the groups observed talked to room guides. 'This lengthened the dwell time in that room, the features remarked upon... and the engagement with stories of the house.' Engagement was also prolonged because visitors were reading room sheets, guide books and replica documents.' In addition, 'Visitors got most from the changes where there was a story and where that story was encountered through a number of rooms in different guises.' (Boyd & Graham, 2010, pp. 4-5, 28).

Some attempts to increase visitor engagement do not appear to have been successful. At the Pitt Rivers Museum old introductory area, 'most spent between 2 and 4 minutes' (James, 2000, p. 9). After redisplay, 'out of 32 accompanied visits only eight people (25%) were recorded as engaging with the introductory display to any extent and even in these cases the words used to describe this suggest a brief engagement, eg. scanned, glanced, briefly, not very interested, skipped it' (James, 2002, p. 6). Commendably, the evaluation report frankly concludes, 'it must be concluded that the introductory display is unsuccessful in its attempt to explain the history of the museum and how it operates. This was not particularly because of its content, but because most visitors failed to notice it at all and those who did, did not engage with it for any length of time.' (James, 2002, p. 9)

In a crowded museum, prolonging engagement may not always be desirable, 'Maintaining visitor attention for long enough, but not too long to impact on visitor flow... is a difficult balance.' (Pontin, 2007, p. 9) **It may be unrealistic to expect most visitors to engage with a large number of exhibits:**

'Observation and interview suggest that visitors can't see and do everything in the gallery. They, therefore, have to choose a personal pocketful of exhibits... It is important to allot your energies to the right exhibits. Interactives, in particular, demand an investment of time and energy to see them through, so the visitor can only take in a few at this high level of commitment. The visitor, therefore, typically checks out an exhibit swiftly to decide whether to pursue it further. During this swift check, the visitor is asking

- Does something interesting/surprising catch my eye?

- Are other people interested in it? (Are there queues, clusters of people, screams, giggles, watchers?).
- Do I care about this subject?' (Fisher & Paynter, 2011, p. 53).

12 Conclusion to Appendix 2, Principal Findings of Summative Evaluation

Generally speaking the evaluations studied paint a rather disheartening picture of visitor disengagement, rather than increasing understanding of what can increase engagement. The corpus of evaluation studies did not allow us to meet all of original objectives for this aspect of the Evaluating Evaluation investigation, which was, in retrospect, ambitious:

‘to identify key and recurrent findings from evaluation studies that have importance and implications for the design and development of displays and exhibitions in museums, galleries, historic buildings and science centres; sub-themes of particular attention will include:

- the contributions made by, and problems that arise from, the use of a variety of interpretive resources including both the traditional and pervasive label and more recent developments such as interactive kiosks and PDAs (Personal Digital Assistants)
- the effect and consequences of mixing digital based interactives with more traditional collections of objects and artefacts
- ways of encouraging interaction between visitors and between visitors and staff;

to identify issues and areas that require further investigation and in particular to clarify where we lack knowledge and understanding of factors that appear to have important bearing upon the success or failure of particular developments.’

In spite of the significant expenditure of time and money on evaluation there remain a great many areas that require further investigation if museums are to have a clearer, empirically-based, understanding of what makes a successful gallery.

From a naïve point of view, this part of the Evaluating Evaluation study was perhaps a search for the ‘golden rules’, or at least ‘higher order principles’ (Griggs 2012) that contribute to the creation of engaging experiences in galleries. We certainly haven’t found those golden rules in evaluation studies. There may still be rules to discover through a better approach to evaluation focused on clearer research questions. However, there may not in fact be many simple golden rules. Participants in the first Evaluating Evaluation colloquium were asked to identify ‘what things do you know (or strongly believe) make for good displays, that create deep visitor engagement?’ The responses were so many and so varied that it suggests there may be no simple answer.

Bibliography

- Arnold, K. (2012). Contribution to Evaluating Evaluation Colloquium 3 December.
- Arts Council England. (2011). *A review of research and literature on museums and libraries*. London: ACE.
- Audience Focus. (2004). *Jameel Gallery of Islamic Art Front-end Evaluation Report*. London: V&A.
- Bishop, G. (2008). *Ice Station Antarctica Summative Evaluation Report*. London: Natural History Museum.
- Bishop, G., & al, e. (2006). *Natural History Museum Balconies Front-End Evaluation #2*. London: NHM.
- BMG Research. (2009). *Research Quality and Scope Assessment Prepared for Renaissance West Midlands*. Birmingham: BMG Research.
- Boyd, N. (2011). *London Street Photography at the Museum of London: Evaluation Report*. London: Museum of London.
- Boyd, N. (2012a). pers comm 6 May 2012.
- Boyd, N. (2012b). Presentation to Evaluating Evaluation Colloquium 3 December.
- Boyd, N., & Graham, J. (2010). *Bringing Properties to Life: The National Trust. Final Report*. unpublished.
- Boyd, N., & James, A. (2009). *Egyptian Galleries: Evaluation Report*. Oxford: Ashmolean Museum.
- Carlyle, S., Glover, T., & Klebanov, A. (2008). *The British Museum: Creating a Methodology for Evaluating a Gallery*. Worcester Polytechnic Institute.
- Carter, J. (2008). *Kelvingrove Text Review*. Glasgow: Glasgow Museums.
- Casson, D. (2012). Presentation at Evaluating Evaluation Colloquium 3 December.
- Clinckemauille, F., Lybarger, A., & Peterson, D. (2010). *Evaluating Visitor Experience in the Department of Coins and Medals at the British Museum*. Worcester Polytechnic Institute.
- Creative Research. (2002). *Summative Evaluation of the British Galleries: Report of Research Findings*. London: V&A.
- Creative Research. (2005). *Darwin Centre Phase II Front End Evaluation Report*. London: Natural History Museum.
- Creative Research. (2008a). *Darwin Centre Phase 2. Key Findings from 3rd Formative Evaluation*. London: NHM.
- Creative Research. (2008b). *Museum of London Case Captions Evaluation Summary Report*. London: Museum of London.
- Culture and Sport Evidence Programme. (2010). *Understanding the impact of engagement in culture and sport: A systematic review of the learning impacts for young people*. London: DCMS
<http://www.culture.gov.uk/images/research/CASE-systematic-review-July10.pdf>.
- Davis, S. (2007). *Exhibit Prototype Testing: Lessons from the Launchpad Project*. London: Science Museum.

- Dawson, E. (2006). *The Ship: the art of climate change: an evaluation*. London: Natural History Museum.
- Dawson, E., & Jensen, E. (2011). Towards a contextual turn in visitor studies: Evaluating visitor segmentation and identity-related motivations. *Visitor Studies*, 14(2), 127-140.
- Dewitt, J. (2012). pers comm 21 March 2012.
- Doeser, J. (2012). Contribution to Evaluating Evaluation Colloquium 3 December.
- East of England Museum Hub. (2008). *Evaluation Toolkit for Museum Practitioners*. Norwich: East of England Museum Hub.
- Featherstone, H. (2012). Contribution to Evaluating Evaluation Colloquium 3 December.
- Fisher, S., & Paynter, F. (2011). *Who Am I? Summative Evaluation 2010 Redevelopment (unpublished)*. London: Science Museum.
- Forrest, R. (2012, June 8). *Evaluation: it's a culture not a report*. Retrieved December 18, 2012, from EVRNN: Visitor experience in museums, galleries, libraries, zoos and botanic gardens: http://evrsig.blogspot.com.au/2012_06_01_archive.html.
- Francis, D. et al. (2011). An evaluation of object-centred approaches to interpretation at the British Museum. In J. Fritsch ed, *Museum Gallery Interpretation and Material Culture* (pp. 153-164). Abingdon: Routledge.
- Frost, S. (2012a). Presentation to Evaluating Evaluation Colloquium 3 December.
- Frost, S. (2012b). pers comm 21 May 2012.
- Fusion. (2008). *The Jameel Gallery of Islamic Middle East Summative Evaluation Report*. London: V&A.
- Fusion. (2011a). *V&A Case Study Evaluation of Future Plan: Medieval & Renaissance and Ceramics Galleries phase I and II Summary of Findings*. London: V&A.
- Fusion. (2011b). *Case Study Evaluation of FuturePlan: Ceramics Galleries Phases I and II Executive Summary*. London: V&A.
- Gammon, B. (nd(a)). *What I've learnt about making successful object centred exhibitions*. Retrieved December 17, 2012, from BenGammon.com: <http://www.bengammon.com/downloads/LessonsForInterpretingObjects.pdf>.
- Gammon, B. (nd(b)). *Planning perfect evaluation of museum exhibits*. Retrieved March 25, 2012, from Ben Gammon: <http://www.bengammon.com/advice.html>.
- Gammon, B., et al (2011). *All about us: Final report prepared for Wellcome Trust/Completion Report (unpublished)*. Bristol: @ Bristol.
- GHK Consulting. (2012). *Review of Informal Science Learning*. London: Wellcome Trust.
- Griggs, S. (2012a). Presentation to Evaluating Evaluation Colloquium 22 February.
- Griggs, S. (2012b). Presentation to Evaluating Evaluation Colloquium 3 December.
- Groves, S. (2012). pers comm 11 December.
- Harknett, S. (2012). Assembling Bodies exhibition evaluation. *UMG/NCCPE Impact and Evaluation Seminar*. Newcastle: UMG/NCCPE.
- Hayward, N, & Burch, A. (2005). *Summative Evaluation: Energy- Fuelling the Future*. London: Science Museum.

- Heath, C.C. & Vom Lehn, D. (2004). Configuring reception: looking at exhibits in museums and galleries. *Theory Culture and Society*, 21(6), 43-65.
- Heritage Lottery Fund. (2008). *Evaluating Your HLF Project*. London: HLF.
- Higgins, P. (2012a). Contribution to Evaluating Evaluation Colloquium 3 December.
- Higgins, P. (2012b). personal communication 3 December 2012.
- Honeybourne, A. (2009). Turner Prize 2008 Reading Room Report. *Tate Papers*.
- Humphrey, T. et al. (2005). *Fostering Active Prolonged Engagement: The art of creating APE exhibits*. Walnut Creek: Left Coast Press.
- Institute for Learning Innovation. (2004). *Formative Evaluation Report for the LACMALab nano Exhibition*. Annapolis, MD: Institute for Learning Innovation.
- Ironside, S. (2009). *Glasgow Museum of Transport (Riverside Project) Text Testing - Families Evaluation Report*. Glasgow: Glasgow Museums.
- Jackson, A. et al. (2006). *Evaluation of the J. Paul Getty Museum's Art Detective Cards Program*. Santa Monica: J Paul Getty Museum.
- James, A. (2000). *Pitt Rivers Museum Evaluation Report [Formative]*. Oxford: Pitt Rivers.
- James, A. (2002). *Pitt Rivers Museum Summative Evaluation Report DCF-Funded Redisplays*. Oxford: Pitt Rivers .
- Jensen, E. (2012). Presentation to Evaluating Evaluation Seminar 22 February.
- King, E. (2012a). pers comm 18 Jul and 3 Sep 2012.
- King, E. (2012b). Presentation to Evaluating Evaluation Colloquium 3 December.
- Kohler, S. (2012). Presentation to Evaluating Evaluation Colloquium 22 February.
- Koutsika, G. (2012). Contribution to Evaluating Evaluation Colloquium 3 December.
- Kuechler, C. (2009). *Evaluation: Clock and Watches Gallery (unpublished powerpoint presentation)*. London: British Museum.
- MacDonald, S. (2012). Contribution to Evaluating Evaluation Colloquium 3 December.
- Mair, V. (2008). *Lives in Motion Exhibition: Summative Evaluation Report*. Glasgow: Glasgow Museum of Transport.
- McManus, P. (2003). *A qualitative account of visitor experiences in the displays, film rooms and study areas of the British Galleries at the V&A*. London: V&A.
- Mead, M. (1995). *Visual Anthropology and the Discipline of Words*. In P. Hockings (Ed.), *Principles of Visual Anthropology (2nd Edition)* (pp. 3-10). London and New York: Mouton de Gruyter.
- MHM. (2005). *Never mind the width, feel the quality*. Manchester: MHM.
- MHM. (2008a). *Proposing Engagement: Visitor Reactions to Church and Emperor: an Ethiopian Crucifixion in Room 3*. London: BM.
- MHM. (2008b). *Touching History: Evaluation of Hands On Desks at the British Museum*. London: BM.
- MHM. (2009a). *A Privileged Insight: An evaluation of Shah 'Abbas: The Remaking of Iran*. London: British Museum.
- MHM. (2009b). *Part of my City, Part of my Heritage: Visitor Research at Kelvingrove Draft 2*. Unpublished.

- MHM. (2009c). *A Privileged Insight: An evaluation of Shah 'Abbas: The Remaking of Iran*. London: British Museum.
- Moussouri, T. (2012). Contribution to Evaluating Evaluation Colloquium 3 December.
- Nomikou, E. (2011). The other side of the coin: audience consultation and the interpretation of numismatic collections. In J. Fritsch, ed, *Museum Gallery Interpretation and Material Culture* (pp. 165-175). Abingdon: Routledge.
- Orna-Ornstein, J. (c1998). Evaluating the Gallery. In Unknown volume, *About the HSBC Money Gallery* (pp. 11-25).
- Pegram, E. (2012a). Presentation to Evaluating Evaluation Colloquium 22 February.
- Pegram, E. (2012b). Contribution to Evaluating Evaluation Colloquium 3 December.
- Playtrain. (2008). *Children and Young People Consultation, Ancient House, Thetford*. Thetford: Ancient House.
- Pontin, K. (2007). *Dino Jaws Summative Evaluation Report*. London: Natural History Museum.
- Pontin, K. (2010). *Using Previous Evaluations for Future Planning*. London: Imperial War Museum.
- RCUK. (2011). *Evaluation: Practical guidelines*. Swindon: Research Councils UK.
- Reussner, E. (2009). *Die Öffnung von Museen für ihr Publikum - Erfolgsfaktoren wirksamer Publikumsforschung*. PhD Dissertation. Berlin: Freie Universität.
- Rodenhurst, K., & al, e. (2008). *Impacts 08 (October 2008) Volunteering for Culture*. Liverpool:
<http://www.liv.ac.uk/impacts08/Papers/Impacts08%282008-10%29VolunteeringForCulture-FINAL.pdf>.
- Science Museum. (nd). *What we've learned about learning from objects*. Retrieved December 17, 2012, from http://www.sciencemuseum.org.uk/about_us/~media/E61EB2E08F6441FC86705D3342634A03.ashx.
- Scotinform. (2003). *Glasgow Museum of Transport Visitor Research Final Report*. Glasgow: Museum of Transport.
- Scotinform. (2008). *Glasgow Museum of Transport 2008 Visitor Survey Research Report*. Edinburgh: Scotinform.
- Serrell, B. (nd c2010). *Paying more attention to Paying Attention*.
<http://caise.insci.org/news/96/51/Paying-More-Attention-to-Paying-Attention>.
- Sinclair, S. (2012). Contribution to Evaluating Evaluation Colloquium 3 December.
- Steiner, K. (2012). pers comm 17 December 2012.
- Stewart, B. (2012). Contribution to Evaluating Evaluation Colloquium 3 December.
- Talbott, F. (2012). Presentation to Evaluating Evaluation Colloquium 3 December.
- Tate Learning. (2010). *Research and Evaluation Strategy Draft 2*.
- Trench, L. (2009). *Gallery text at the V&A: A ten-point guide*. London: V&A.
- Urquhart, G. (2001). *The visitor profile of the Glasgow Museum of Transport Interim Report 15 July 2001*. Glasgow: Glasgow Museums.

- vom Lehn, D, Heath C. C. and Hindmarsh, J. (2001). Exhibiting Interaction: Conduct and Collaboration in Museums and Galleries Interaction. *Symbolic Interaction*, 24(2), 189-217.
- Walmsley, B. (nd 2011?). *Why people go to the theatre: a qualitative study of audience motivation*.
http://www.eventsandfestivalsresearch.com/files/proceedings/WALMSLEY_FINAL.pdf.
- Watl, C. (2006). Museums for visitors: Audience development. *Intercom*.
<http://www.intercom.museum/documents/1-4Watl.pdf>.
- Wellcome Trust. (2012). *Informal Science Learning Review: Reflections from the Wellcome Trust*. London: Wellcome Trust.
- Wilkins, T. (2012). Presentation at Evaluating Evaluation Colloquium 3 December.