

Original citation:

Jensen, Eric (2015) Highlighting the value of impact evaluation : enhancing informal science learning and public engagement theory and practice. Journal of Science Communication, 14 (3).

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/92998>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0) license and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

A note on versions:

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Highlighting the value of impact evaluation: Enhancing informal science learning and public engagement theory and practice

JCOM commentary

Dr Eric Jensen
University of Warwick

Abstract

King et al. (2015) argue that 'emphasis on impact is obfuscating the valuable role of evaluation' in informal science learning and public engagement (p. 1). The article touches on a number of important issues pertaining to the role of evaluation, informal learning, science communication and public engagement practice. In this critical response essay, I highlight the article's tendency to construct a straw man version of 'impact evaluation' that is impossible to achieve, while exaggerating the value of simple forms of feedback-based evaluation. I also identify a problematic tendency, evident in the article, to view the role of impact evaluation in advocacy terms rather than as a means of improving practice. I go through the evaluation example presented in the article to highlight alternative, impact-oriented evaluation strategies, which would have more appropriately addressed the targeted outcomes than the methods used by King et al. (2015). I conclude that impact evaluation can be much more widely deployed to deliver essential practical insights for informal learning and public engagement practitioners.

Highlighting the value of impact evaluation: Enhancing informal science learning and public engagement theory and practice

King et al. (2015) argue that ‘emphasis on impact is obfuscating the valuable role of evaluation’ in informal science learning and public engagement (p. 1). The article touches on a number of important issues pertaining to the role of evaluation, informal learning, science communication and public engagement practice. In this critical response essay, I highlight the article’s tendency to construct a straw man version of ‘impact evaluation’ that is impossible to achieve, while exaggerating the value of simple forms of feedback-based evaluation. I also identify a problematic tendency, evident in the article, to view the role of impact evaluation in advocacy terms rather than as a means of improving practice. I go through the evaluation example presented in the article to highlight alternative, impact-oriented evaluation strategies, which would have more appropriately addressed the targeted outcomes than the methods used by King et al. (2015). I conclude that impact evaluation can be much more widely deployed to deliver essential practical insights for informal learning and public engagement practitioners.

Problems of definition: Setting the impact evaluation bar

King et al. (2015) draw a distinction between ‘evaluation studies’ that ‘focus on immediate questions of what and how something works’ on the one hand and ‘impact analyses’ on the other hand. This is a problematic distinction. In fact, good “impact evaluations are concerned with establishing what works and why” (Owen & Rogers, 1999, p. 256).

In their article, King et al. conjure an intimidating image of ‘impact analysis’ as “longitudinal; large-scale; involving control populations; with both quantitative and qualitative elements” (p. 2). They go even further by indicating that impact evaluation requires using randomized control trials of the type employed in pharmaceutical research:

To measure the impact of new drugs, the field of medical research use research methodologies that require double blind tests, and randomised control trials (RCTs). These approaches involve large numbers of research subjects and follow standardised routines which enable experiments to be replicated elsewhere, all of which allow researchers to report findings with high degrees of statistical validity and reliability. Conducting large scale RCTs in informal learning settings, however, is highly challenging due to the difficulty of recruiting the large numbers of visitors required for statistical analysis, and moreover negotiating the ethical and practical dilemmas of ensuring that each research subject has a similar type of experience whether they participate in the intervention or the control. (p. 3)

This sets a ludicrously high bar for impact evaluation of informal science learning and public engagement activities. Moreover, a randomized control trial (used to evaluate whether a new pharmaceutical product performs better than a placebo) is a clearly inappropriate model for use in an informal science learning or public engagement context. Indeed, it is inappropriate for almost all social contexts in which impact evaluation is used (i.e. outside of laboratory settings).

This kind of imposition of the conventional scientific evidence hierarchy on impact evaluation in real-world contexts such as education has been widely criticized for these and other reasons (e.g. Chatterji, 2009). It is beyond the scope of this paper to explain in detail, but I would question whether control groups are at all needed or useful for impact evaluation in informal learning or public engagement settings (Wagoner & Jensen, 2015). There is also no reason to assume that impact evaluation must be 'large-scale' or include 'both quantitative and qualitative elements'. Longitudinal research- in the most limited sense of including more than one data collection point with the same individual- can be very important, but this expectation is perfectly feasible in many informal learning and public engagement contexts.

Defining the role of 'impact evaluation' is also important. Davies and Heath (2013) highlight the common pattern that 'the requirement for good news forms the framework in which evaluation takes place' (p. 23). This same 'good news' bias permeates King et al.'s article, as they describe their task in the evaluation they present as "*documenting the success* of a museum-led teacher professional development programme" (emphasis added; p. 2). This a priori of assumption of success is simply a problematic starting point for impact evaluation. It assumes success and sets about trying to prove it. Indeed, this kind of approach is evident throughout King et al.'s (2015) account of their evaluation. This is not research, nor evaluation; it is in the realm of advocacy and campaigning, where evidence is seen as one rhetorical tool amongst others to support a previously established point of view. As a general rule, the point of evaluation should *not* be to prove impact; it should be to evaluate the effects, if any, of a given intervention. All kinds of evaluation results, whether positive or negative, can be useful for enhancing practice.

Conflating 'it hasn't been done' with 'it can't be done'

King et al. rightly point out that much impact evaluation, often conducted by museum consultancies or market research agencies, has been poorly executed (also see Davies & Heath, 2013). Indeed, I have previously pointed out the general pattern that within science centres and museums "industry standard' visitor surveys and evaluation procedures at such institutions offer a virtual catalogue of basic errors and poor practice in survey design, sampling and analysis' (Jensen, 2014b, p. 1).

Poor quality evaluation has been feeding dodgy data and specious conclusions into the science communication system for years. Science communication institutions are generally uncritical consumers (and producers) of evaluation research, quick to believe that measuring complex outcomes can be incredibly simple. Want to know whether a child has learned a lot about science after her day at the science museum? Easy! Just ask her: 'Did you learn during your visit to the science museum today?': Yes or No?. Think I am exaggerating the problem? London's vaunted Science Museum has its own audience research team, and yet its internal guidance for evaluation includes the following flawed survey item, 'To what extent do you agree or disagree with the following statements?' (strongly agree to strongly disagree): 'I

have learnt something new today' (National Museum of Science & Industry, 2009). Of course, measuring science learning, attitudes and other key outcome variables is not actually this simple. When our hypothetical child above says 'yes' to the self-reported learning question, she is most likely telling the institution what it wants to hear. This question imposes the unrealistic expectation that respondents can accurately assess their pre-visit science knowledge, identify any gains or losses that occurred during the visit and correctly self-report their conclusions on a five-point scale. Actually measuring learning requires (at minimum) direct measurement of visitors' thinking or attitudes before and after the intervention. (Jensen, 2014b, p. 2)

However, this widespread problem of poor quality impact evaluation does not mean that it is an impossibly difficult task.

It's (practically) impossible! The myth of 'unevaluatable' impact

An excuse for a lack of robust evaluation evidence that I sometimes encounter within science communication institutions is that evaluating impact in informal learning and engagement contexts is effectively impossible. In making this point, King et al. (2015) construct a straw man argument by setting an unreasonable standard for impact evaluation: '*Proving the causal effect of one particular intervention (over and above other experiences)* and thereafter claiming impact is, therefore, extremely difficult' (p. 2). They also later state that "collecting the data to confidently claim changes in for, example, teacher practice and thereafter student learning (levels 4 and 5) was acknowledged to be impossible" (p. 5-6). Similarly, Trench (2014) alludes to this common excuse of near impossibility with his suggestion that it 'may not be knowable with any certainty' whether a 'major science centre' or 'government programme targetting science awareness' are creating any impacts.

While impact evaluation can be difficult in many contexts, the way it operates is straightforward conceptually: You try to isolate the effect of an intervention (for example, by measuring a child's thinking about an idea before and after an intervention). The social sciences have invested decades of effort into developing the tools to validly research phenomena relating to attitudes, interests, motivations, thought processes and emotions, outcomes which are the subject of most informal learning and engagement goals. There is no achievable science communication goal I have yet encountered (in 14 years of science communication research) that is impervious to robust measurement using social scientific methods. For example, if I know how the same person understands the concept of biodiversity before a zoo visit and after a zoo visit, then I can identify the 'effect' of the zoo visit period on this outcome (Moss, Jensen & Gusset, 2015). This is a straightforward procedure that has been conducted by over 30 zoos and aquariums all over the world without external funding. Qualitative data or multiple data collection time points within the informal learning or engagement experience can further close the inferential gap to be able to robustly attribute impact (whether positive or negative) to an experience.

The value of impact evaluation

Impact evaluation should not be placed on the shelf as an ivory tower ideal, only to be dusted off on rare occasions when an academic comes around with mounds of spare time and resources.

Good impact evaluation requires upstream planning and clear objectives from practitioners. Moreover, results should inform science communication practice. It also requires training (either externally provided or self-taught) in relevant social scientific research methods (e.g. survey design). (Jensen, 2014b, p. 3)

Impact evaluation can be a crucial mechanism for avoiding the risk of unforeseen negative outcomes, which could otherwise go undetected by practitioners for decades. We all know that any communication, learning or engagement activity can miss the mark: We can say things in a way we did not intend, misjudge our audience or fail to communicate effectively in a myriad other ways. We all know that the consequences of failures of this kind can be more serious than merely not getting our message across. If we appear insensitive to the needs of those we are speaking to or otherwise communicate our ideas ineffectively, we risk engendering feelings of resentment or frustration in our audience, making them dislike us or decide that we aren't worth their time. All of these risks apply to informal learning and engagement institutions. If their activities are ineffectively conducted, the consequence can be net negative impacts on audiences. Given the seriousness of such outcomes, it is essential that evaluations allow for the possibility of negative impacts. That way, any negative impacts can be identified and necessary changes implemented immediately.

In a well-functioning informal learning and engagement institution, impact evaluation would already be integrated into practice to ensure that evidence is feeding into on-going improvements for better audience outcomes. 'Working in concert, quality impact evaluation and critical self-reflection by practitioners could be used to counteract stagnation and systemic failures in science communication practice' (Jensen, 2014b, p. 3). I would argue that the 'demands for impact' that King et al. wish to roll back are becoming increasingly difficult to ignore because informal learning and engagement practice has so long eschewed a robust evidence-based approach to long-term improvement. Impact measurement is now being sought by funders rightly wanting to know whether their investment is making a positive difference for audiences. At this relatively late stage in the development of informal learning practice, funders may well be imposing unrealistic expectations that deserve to be challenged. However, informal learning and engagement institutions that have long expected everyone to take their positive impact on faith should accept a share of the blame for this situation. I would argue that the present emphasis on impact is an opportunity to negotiate a more scientific approach to informal learning and engagement, grounded in robust evaluation of what engagement practices are actually achieving (or failing to achieve) *and why*.

Impact evaluation with limited resources

I am sympathetic to the argument that 'for many informal learning institutions providing evidence of change in knowledge, skills or behaviour is extremely difficult given the narrow constraints of available budget, staff and methodological expertise within which such organisations operate' (King et al.,

2015, p. 1). Limitations in available budget, staff and methodological expertise are real problems for many informal learning and public engagement organisations (Jensen, 2014b), undermining their ability to use robust evaluation methods. Indeed, marketers, educators and managers working in informal learning and public engagement organisations are busy people. Most of them have not been trained in the social scientific research methods required to produce valid visitor evaluations or to be a critical consumer of visitor research and impact evaluations conducted by others. This undoubtedly helps to explain the ubiquity of very poor quality visitor research and evaluation in this field, including in much of the published empirical museum and visitor studies literature on informal learning audiences. Below I discuss two main options for overcoming these problems: Improving methodological knowledge within informal learning institutions and using technology-enhanced methods of evaluation.

Some methodological understanding would certainly be beneficial for informal learning and public engagement staff, who will encounter evaluation evidence over the course of their careers. It is important to be a savvy consumer of evaluation research, able to identify and avoid common limitations such as some of the issues identified in this paper. Today however, there is less of a need for practitioners to develop the practical knowledge and skills to be able to implement high quality evaluations for themselves. This is because recent improvements in open source technology bring good quality evaluation within easy reach of many more informal learning and public engagement organisations.

For many, if not most, informal learning organisations, technology-enhanced evaluation could be a real solution for embedding robust evidence within the fabric of informal learning and engagement institutions. Automated evaluation tools enable answers to questions such as, what proportion of visitors are satisfied with their experiences? And, what factors are affecting the quantity and type of impact on visitors? Automated methods of evaluation can eliminate the need for on-going costs and expensive external consultants in order to gain evaluation evidence. New technologies enable the design of evaluation systems that can be fully automated after an initial customization and set-up. Using these technologies, a one-time infusion of expertise can create a system used by practitioners without any skills in social scientific analysis. Recently developed options using open source technology include:

- A system of evaluation that is integrated into an institution's visitor smartphone app to automatically gather, analyse and display for the institution evaluation results such as visitor timing and tracking information and micro-survey feedback through the app. (qualia.org.uk)
- An automated system that identifies indicators of 'quality of experience' in social media messages posted by visitors, using categories developed through online ethnographic research with people tweeting about public engagement experiences. (culturesmile.org)
- A city-wide 'what's on' smartphone app, which uses in-app micro-surveys based on shared metrics across participating institutions to develop evaluation knowledge showing the flow and engagement of audiences across institutions. (artory.co.uk)

In each of these cases, data collection and analysis are completely automated, with visualisations of the results displayed for institutions using the systems in real time. This means that institutions can immediately act on incoming evaluation results, rather than waiting for cumbersome processes of data collection, data entry and analysis to take place.

Of course, any research approach has strengths and limitations. Automated evaluation methods are not suitable for every evaluation challenge. However, greater adoption of these technologies could raise the minimum standard of evaluation in the sector, and provide institutions with a finger on the pulse of their audiences. Because these systems have been built using robust open source software development, the systems can be easily adopted by organizations and interface with other commonly used software such as Eventbrite. The result is greater availability of higher quality evaluation evidence, while organisations conserve resources by replacing existing consultancy costs or diverted staff time. Ideally, if the resource burden of on-going visitor evaluation and market research could be removed, the sector would then be able to focus on strategic investment in in-depth rigorous research on aspects of visitor engagement that require particular attention (for example, non-visitors) (see Dawson & Jensen, 2011; Jensen, Dawson & Falk, 2011).

Evaluating the ‘Talk Science’ Evaluation: Problems and alternative strategies

Evaluation is incredibly important for informal learning and public engagement practice: Without it, staff are using unreliable ‘gut instincts’ to intuit whether effective experiences are being provided. High quality evaluation and visitor research that is skillfully conducted and effectively shared can provide a basis for practitioners to discover what aspects of an experience are working, in what ways, with which audiences and why.

Against this backdrop, I agree that the kind of simple evaluation demonstrated in this article for the ‘Talk Science’ project can be valuable. Feedback forms can establish whether an enjoyable experience is being delivered and whether there are any major concerns from participants. However, King et al. overstate what can be gained from such basic feedback evaluation without an impact measurement element. The article applies a framework from Thomas Guskey that identifies five outcomes for an effective teacher development experience:

- Outcome 1. Participating teachers have a positive learning experience.
- Outcome 2. Participating teachers gain new knowledge and skills.
- Outcome 3. Change is supported in the participating teacher’s school.
- Outcome 4. The new teaching approaches are implemented.
- Outcome 5. Gains are visible in student learning.

Notably, only part of one of these outcomes has been effectively evaluated using the approach presented in the article. Gathering basic feedback as was done in this case does enable practitioners to establish whether an experience was ‘positive’ (part of Outcome 1 above). However, the remaining outcomes were not

robustly evaluated. I will now take each of these outcomes one at a time to show how they could have been measured.

Outcome 1. Participating teachers have a positive learning experience.

For this outcome, you would need to evaluate whether the teachers had a 'learning experience', and if so, what kind of learning developed? (e.g. it could be positive or negative learning, as they may have learned good or bad ideas and habits). One way to do this would be to interview teachers (using semi-structured qualitative interviews), asking them to talk about what the CPD experience was like from their perspective. A robust qualitative analysis of what they say would enable valid conclusions about the nature of learning impacts (of all kinds) that emerged from the intervention, as well as the process through which that learning developed.

Outcome 2. Participating teachers gain new knowledge and skills.

This outcome requires impact evaluation to be able to assess what kind of changes occurred from pre- to post-teacher CPD event. One option here would be to use a qualitative survey to conduct pre- and post-measurement of teacher's thinking about key issues addressed by the CPD intervention. For example, I conducted an extremely low budget impact evaluation of a training programme in Mauritius designed to develop local hotel guides' understanding of the coral reefs nearby so they could deliver better conservation education to tourists. A printed questionnaire was used for both the pre- and post-training learning measurements. Delegates were initially surveyed on the first day of the training, when they were asked open-ended (qualitative) impact evaluation questions and close-ended (quantitative) questions about their demographic characteristics. The post-training questionnaire included feedback items as well as the second iteration of open-ended impact evaluation questions. Data were entered into a spreadsheet for analysis. Pre- and post-training responses were compared to assess the impact of the training. As can be seen from this example, evaluating the impact of similar projects, although a time-consuming process, is not complex and greatly aids the development of a conservation education course.

The approach to analysing the evaluation questionnaire data was first to consider the learning impacts from the training outcome-by-outcome, then to analyse the individual trajectories for each delegate on a case-by-case basis. Analysis was primarily qualitative. Data were analysed following inductive techniques and procedures to identify patterns. The analysis followed an adapted form of 'grounded methodology' for qualitative data analysis (see Jensen & Holliman, 2009), building up an understanding of the impacts of the training activity inductively from the data. There were also small quantitative content analysis aspects included in the analysis for the sake of clarity.

Unlike in the example presented by King et al. (2015), my evaluation explicitly sought out negative feedback in addition to the positive feedback:

The 'worst' part of the course identified by two of the delegates was the use of scientific jargon. Two others stated their dislike of the octopus dissection activity. Aside from these concerns, only Andre (Male) voiced concern over the introduction to the course, 'Not [the] worst, but it takes time to understand at the beginning [of the training course].' (Jensen, in press)



In addition to feedback, I directly measured impact using qualitative methods. There was evidence of transformations in delegates' thinking from pre- to post-training. The main patterns of positive change evident in the thought-listing data (e.g. qualitative responses to 'What comes to mind when you think of a "coral reef"?') were: (1) increased elaboration in delegates' understanding of habitat the coral reef provides for animals and (2) the benefits the reef brings to the island and its human population and (3) greater focus on conservation of coral reef environments. This kind of broad elaboration in delegates' thinking can be seen in the example of Henri's pre- and post-training responses to the question 'What comes to mind when you think of a "coral reef"?' below.

Pre-Training (Henri)			
Polyps	Sand	Important for Fish	The Great Barrier Reef

While Henri identified relevant thoughts above, his thinking clearly became sharpened and more elaborated following the training course.

Post-Training (Henri)			
Zooxanthellae and all the living animals we may find in coral reefs compounds	Need [to] protect them because they [are being damaged]	Without coral reefs, human beings life will definitely change; no more food for the sea	So beautiful and amazing--such lovely colours down there

In addition to the thought-listing impact measure, following an established method (e.g. Jensen, 2014a) I included a survey question asking delegates to draw a coral reef with all the plants and animals that live there. They were asked to label everything to aid the analysis. The drawings were scanned, then individual elements were extracted and analyzed systematically. Below is just one example of a 'data extract' from this kind of impact evaluation:

	Pre-training	Post-training
Henri		

This relatively low-budget example of impact evaluation using primarily qualitative methods highlights that direct, robust measurement of intended knowledge outcomes is in fact feasible. Of course, methodological knowledge and skills were required to design and implement the evaluation, so it is not an easy solution for most practitioners. This is an example of an evaluation that could not be done with automated methods, as the project took place in a context with low availability of technology and using an outcome measure (annotated drawings) that could not be analyzed automatically. With teachers in a developed country however, it would certainly be feasible to implement an automated evaluation using online questionnaires.

Outcome 3. Change is supported in the participating teacher's school.

This is a rather vague outcome as stated here. If it is referring to the degree to which a teacher is empowered by the structure at his or her school to make changes based on the new information or skills he or she had developed, this could be measured straightforwardly, either with a closed-ended survey question asking the teacher to rate the extent to which he or she felt empowered to make changes generally, and then specifically in this case. Or, an open-ended survey question or qualitative interview could explore the teacher's perception of the management structure or other factors affecting his or her ability to make changes. This would be a way of measuring whether the conditions were in place for further impact to develop once the teachers returned to their school.

Outcome 4. The new teaching approaches are implemented.

Obviously, the ideal for this outcome would be to directly and systematically observe relevant teacher practices before and after the CPD intervention. That should have been possible in the case presented by King et al. (2015) given there was a dedicated researcher on the project. However, it would normally require too much time and resources to do this. So, a useful proxy measure could be to ask an open-ended question where teachers describe the relevant aspect of their teaching practices before and after the CPD event (and again some time later, e.g. 3 or 6 months later). Robust analysis of these data could reveal whether there had been any change over this time period, and if yes, what kind of change had occurred.

In contrast, the approach presented by King et al. (2015) is an example of problematic evaluation design. They report '90% of respondents commented that they had actively disseminated new knowledge and skills to their colleagues back at school' (p. 7). Even assuming these self-reports are accurate, a problem with this measure is that it does not account for the specific 'knowledge and skills' that were disseminated. If teachers had misunderstood or misconstrued key messages, then they may have disseminated these misunderstandings amongst their colleagues, thereby spread poor practice. Given that the evaluation method described in King et al. (2015) cannot address whether or not such negative impact took place, its value for shedding any light on this outcome is questionable.

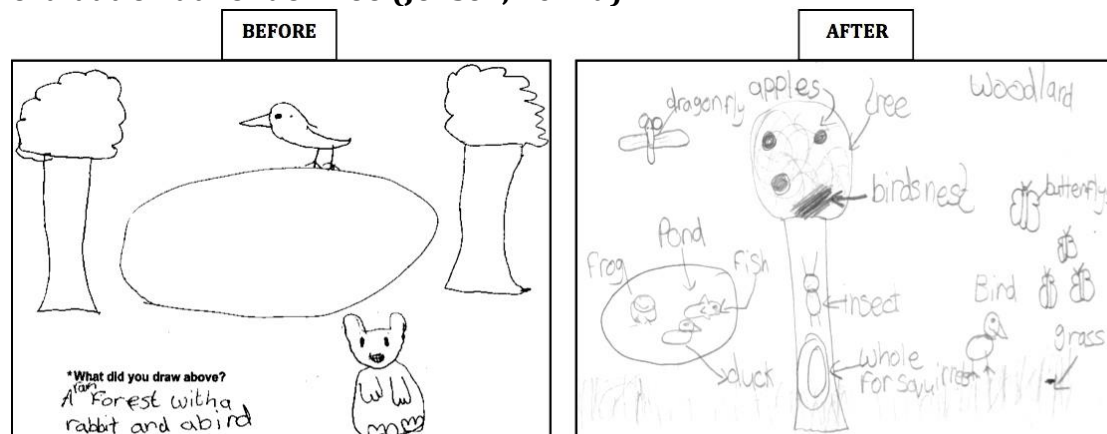
Outcome 5. Gains are visible in student learning.

King et al. (2015) aver that 'Finding evidence of Guskey's level 5 is [...] particularly hard to demonstrate' (p. 8). While it is not reasonable to expect

every teacher CPD initiative to conduct impact evaluation on the effects of the CPD for children, it is certainly possible to do so. Teachers signing up to the CPD event could be sent a package of questionnaires to hand out to their pupils with instructions. These completed questionnaires are handed in when the teacher arrives at the CPD event (or before, using a self-addressed pre-stamped envelope). Teachers are then given a matching post-intervention questionnaire to give out at a specified point after the CPD event (e.g. 2 or 4 weeks later) to see if the new practices taught in the CPD event were evident in pupils' assessment of the teaching and learning activities in their class and their understanding of the topics targeted by the intervention. If differences are found, they could be analysed to see if they are in the direction that would be expected based on the CPD event content. This is not conclusive impact evidence, but it would provide some way of testing how the CPD event's effects are (not) making their way into pupils' experiences.

As a general principle, to measure children's attitudes or learning, you must gather data directly from them using appropriate language and good survey design techniques. For example, after pilot testing different impact evaluation question options (Wagoner & Jensen, 2010) for evaluating children's learning at London Zoo, I used this question: 'Please draw your favourite wildlife habitat and all the plants and animals that live there (put names and labels on everything)'. This item yielded annotated drawings from questionnaire data gathered from pupils by their teachers before and after their visit to London Zoo. These annotated drawings were then subjected to rigorous content analysis to quantify the impact patterns (and variables that predicted impact), as well as a conventional qualitative analysis to elaborate understanding of the details of what was happening with children's understanding of wildlife habitats over the course of a zoo visit (Jensen, 2014a).

Figure 1: Pre- and post-visit drawings for same child from impact evaluation at London Zoo (Jensen, 2014a)



In contrast to the above approach, on this variable, King et al. (2015) give the disclaimer about not being able to 'confidently claim changes in for, example, teacher practice and thereafter student' (p. 6). However, they then claim that 'teacher self-reports...provided the team with a source of insight and suggestions for improvement' (p. 6), presenting teacher self-report data as evidence of 'possible' impact, despite the initial disclaimer:

Nonetheless, the evaluation findings offered the team an insight into the areas of student knowledge or skills that the course potentially enhanced. For example, the following quotes highlight student acquisition of discussion skills, and increases in student engagement with science:

It's improved their speaking skills. Quite a lot of them are used to giving one-word answers whereas now they're expanding upon what they're talking about. They've become a little bit more eloquent.

Science teacher, phone interview

The kids were engaged, they were excited. I just think they were learning, they were talking, they were asking questions which they wanted to know.

Science Teacher focus group interview (King et al. 2015, p. 8)

Here we can see King et al. (2015) indicating that they addressed this outcome by asking teachers about the effects on their pupils. This widely used approach of 'other report' by teachers or parents is clearly flawed. Teachers are not psychics: They are not going to be able to accurately assess effects on the full range of their pupils without implementing their own systematic evaluations. It is very likely that teachers are merely cherry picking positive stories to tell the Science Museum about their programme.

Conclusion

King et al. (2015) are right to point out that it requires methodological expertise and resources to do good impact evaluation. Clearly current evaluation practices are problematic. King et al. (2015) frame impact evaluation as an impossibly difficult task. Yet, in the broader world of programme evaluation (outside of informal learning, science communication and public engagement), impact evaluation is arguably the most used form of evaluation (Owen & Rogers, 1999). Developing appropriate skills and knowledge to use impact evaluation to inform informal learning and engagement practice should not be an insurmountable barrier. At one point, I advocated training for staff to redress the problems identified in this essay. I still think a degree of training in research methodology would be beneficial for practitioners, if nothing else so that practitioners can be savvy, critical consumers of research conducted by consultancies and academics. However, recent improvements in technology open up new options for implementing systems that provide on-going evaluation insights on an automated basis with a one-time infusion of expertise at the survey design stage (e.g. see qualia.org.uk or artory.co.uk). Given the logistical challenges involved in developing high quality, practical research methods training for practitioners whose primary responsibilities lie elsewhere, such technology may be a big part of the answer to the challenge of implementing robust evaluation in informal learning and engagement institutions. On-going evaluation systems would allow informal learning and engagement organisations to be much better attuned to their audiences' needs, less reliant on speculation about which interventions are effective and therefore more likely to deliver positive impacts. Using robust social scientific evidence in the form of

evaluation or audience research to ensure success should be viewed as a basic necessity across the sector.

References

- Chatterji, M. (2009). Enhancing impact evidence on how global education initiatives work: Theory, epistemological foundations, and principles for applying multiphase, mixed method designs. In K. E. Ryan, & J. B. Cousins (Eds.), *The SAGE International Handbook of Educational Evaluation*. (pp. 97-113). Thousand Oaks, CA: SAGE.
- Davies, M., & Heath, C. (2013). *Evaluating Evaluation: Increasing the Impact of Summative Evaluation in Museums and Galleries*. Last accessed at: http://visitors.org.uk/files/Evaluating_Evaluation_Maurice_Davies.pdf
- Dawson, E. and Jensen, E. (2011). Towards a 'contextual turn' in visitor research: Evaluating visitor segmentation and identity-related motivations. *Visitor Studies*, 14(2): 127-140.
- Humphreys, L. & Jensen, E. (in press). Mobilising tourism for conservation education. In E. Jensen, *Making the most of conservation education: Research, principles and practice*. Cambridge: Cambridge University Press.
- Jensen, E., Dawson, E. and Falk, J. (2011). Dialogue and synthesis: Developing consensus in visitor research methodology. *Visitor Studies*, 14(2): 158-161.
- Jensen, E. (2011) 'Evaluate impact of communication'. *Nature*, 469(162).
- Jensen, E. (2014a). 'Evaluating children's conservation biology learning at the zoo'. *Conservation Biology*, 28(4): 1004-1011.
- Jensen, E. (2014b). 'The problems with science communication evaluation'. *Journal of Science Communication*, 1(2014)C04. Accessed at http://jcom.sissa.it/archive/13/01/JCOM_1301_2014_C04/JCOM_1301_2014_C04.pdf.
- King, H., Steiner, K., Hobson, M., Robinson A. and Clipson, H. (2015). 'Highlighting the value of evidence-based evaluation: pushing back on demands for 'impact''. *JCOM* 14(02), A01.
- Moss, A., Jensen, E. & Gusset, M. (2015). 'Evaluating the contribution of zoos and aquariums to Aichi Biodiversity Target 1'. *Conservation Biology*, 29(2): 537-544.
- Owen, J. & Rogers, P. J. (1999). Impact evaluation. In J. M. Owen, & P. J. Rogers (Eds.), *Program evaluation*. (pp. 263-308). London: SAGE.
- Trench, B. 'Do we know the value of what we are doing?'. *Journal of Science Communication*, 1(2014)E. Accessed at http://jcom.sissa.it/sites/default/files/documents/JCOM_1301_2014_E.pdf.

Wagoner, B. & Jensen, E. (2010). 'Science learning at the zoo: Evaluating children's developing understanding of animals and their habitats'. *Psychology & Society*, 3(1): 65-76.

Wagoner, B. & Jensen, E. (2015). 'Microgenetic evaluation: Studying learning in motion', In G. Marsico, R. Ruggieri & S. Salvatore (Eds.), *The Yearbook of Idiographic Science. Volume 6: Reflexivity and Change in Psychology*. Charlotte, N.C.: Information Age Publishing.