

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

Eric Jensen

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 exemplified in the article. 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 addressed the targeted outcomes more appropriately 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.

Keywords

Science centres and museums

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 exemplified in the article. 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 and Rogers, 1999, p. 256].

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

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 clearly an 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). The imposition of this conventional 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 and 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 — is essential, but this expectation is perfectly feasible in many cases within 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 framework 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” (p. 2). This 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. 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 results, whether positive, negative or no impact proof, are 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. Indeed, I have previously pointed out 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]. Indeed, I used an example from London’s Science Museum to evidence this point:

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 does not mean that it is an impossibly difficult task.

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

Another excuse is that evaluating impact in informal learning and engagement contexts is impossible. Here again, 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 “measuring confidently claim changes in for, example, teacher practice and thereafter student learning (levels 4 and 5) was acknowledged to be impossible” (p. 6). Trench [2014] alludes to this 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, 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 science communication goals. There is no achievable science communication goal I have yet encountered (in 13 years of science communication research) that is impervious to robust measurement using social scientific methods. 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 and 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 form of communication can miss the mark: we can say things in a way we didn't intend, misjudge our audience or fail to communicate effectively in a myriad of 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 science communication. If it is 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 system, impact evaluation would already be integrated into practice to ensure that evidence is feeding graduate improvement. As I have argued, '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]. Because informal learning and engagement practice has so long eschewed an evidence-based approach, the 'demands for impact' that King et al. wish to roll back are being imposed by funders and others rightly wanting to know whether their investment is well-placed. At this relatively late stage in the development 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. It has long struck me as ironic that those institutions and individuals touting the value of the scientific method for society (e.g. the London Science Museum) are so reluctant to prioritize a robust evidence-based APPROACH in their own practices. 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 is a real problem for many informal learning and public engagement organisations, a concern which I have highlighted previously [Jensen, 2014b].

The fact that the London Science Museum focuses its efforts on formative evaluation is entirely appropriate. However, it has had an entire 'Audience and Advocacy team' for many years now: it would be reasonable to expect that at some point a full-scale impact evaluation study could have been conducted. Moreover, the article states that a 'dedicated Audience Researcher was recruited' for this project' (p. 5). This indicates that the problem in the case presented in the article is not a 'lack of resources', but rather institutional failure in the recruitment and selection process resulting in a lack of appropriate social scientific methodological expertise within the institution. Of course, existing staff members are capable of learning new skills and gaining new capacities, so a lack of methodological expertise does not have to be a permanent condition.

**Solutions to
improve
on-going
evaluation
capacity**

Technology is one answer

While some methodological understanding would still be beneficial for informal learning and public engagement staff, recent improvements in open source technology brings good quality evaluation within easy reach of many more informal learning and public engagement organisations.

Marketers, educators and managers working in informal learning and public engagement organisations are busy people. Most of them have not been trained in social scientific research methods required to produce valid visitor evaluations or be a critical consumer of visitor research and impact evaluations conducted by others. This no doubt helps to explain the ubiquity of very poor quality visitor research and evaluation in this field, including in much of the published empirical visitor studies literature. A couple of years back, I tried one option to help the situation by offering a series of training seminars on impact evaluation of public engagement and informal learning with support from the Wellcome Trust and the British Science Association: <http://www2.warwick.ac.uk/fac/soc/sociology/news/scd/publicengagement>. I have also led other training courses on evaluation

and visitor studies methods at conferences and other events for practitioners. While people came to these sessions, I do not think training per se is a feasible solution for the sector as a whole, unless the ability to design, conduct and analyze visitor evaluations comes to be viewed as a core skill for audience-facing staff. This is because it takes time to learn how to design and then to actually conduct valid data collection and analysis, and many informal learning and engagement professionals are just too pressed with other priorities to undergo such extensive training.

Therefore, I have come to the conclusion that the best contribution I can make for the sector is to provide evaluation templates and systems that do not require expertise in social scientific research methods. Firstly, I have published on my website examples of impact evaluation survey questions and forms I have designed for past projects that I am happy for informal learning and engagement staff to re-use. For example, on page 36 of this report are impact survey forms and questions that could be adapted for use by practitioners:
https://www.academia.edu/6285529/A_Global_Evaluation_of_Biodiversity_Literacy_in_Zoo_and_Aquarium_Visitors.

Most recently, I have explored an alternative that could be a real solution for informal learning and engagement staff interested in embedding evaluation and visitor research into the fabric of their institutions. What proportion of visitors are satisfied with their experiences? What factors are affecting the quantity and type of impact you are having on visitors? These are just a few of the questions that can be addressed using automated evaluation and feedback tools. Automated methods of evaluation can be used to gather and analyze large amounts of data from multiple sources. The results they produce can provide valid insights about what aspects of informal learning and engagement experiences are yielding positive or negative results.

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. I explored this possibility over the last few years through the Qualia project (funded by a UK government's Digital R and D Fund for the Arts): qualia.org.uk; the SMILE project (<http://www2.warwick.ac.uk/fac/soc/sociology/staff/academicstaff/jensen/ericjensen/smile>), Artory (artory.co.uk) and other research. The goal was to use a one-time investment of public money to build high quality open source evaluation systems that could be used by cultural institutions with their visitors to gather evaluation data automatically, with a bare minimum of customization required to deliver real-time results. These systems are complete and ready for use to evaluate informal learning and engagement experiences.

Technologies that could be used for evaluation include:

1. *Visitor app* — Native iPhone and Android app includes schedule function for talks and other events (automatic feedback questions after events), social media function for Twitter and Facebook posts through the app, map, information, automatic feedback request after visit either in app or by short linked web survey. While this app builds on technical infrastructure that built

and tested for the Qualia project, it appears to be a science communication organisation's own visitor app, with only its branding on it.

2. *Visitor insight dashboard* with real-time results of automated analyses feeding in. Automatic alerts can be sent to relevant personal when feedback dips below minimum threshold, etc.
3. *Information kiosk touchscreens* with event information, site map and feedback features that automatically feed into the analytics.
4. Automated *sentiment analysis* of the content of social media data pertaining to experiences at the institution.

I subjected these different technologies to rigorous validity and reliability testing as part of the Qualia project, and in the end I do think it is a workable system. It has been built using open source software development, so the system can be easily adopted by organisations (saving enormous sums of money across the sector by replacing existing consultancy costs and offering better quality, real-time visitor evaluation results). If this kind of system is widely adopted over time, then it would be feasible to build up sector-wide data on impacts. This could also be achieved using conventional social scientific methods, but that would require a much larger on-going investment of resources. Ideally, if the cost burden of on-going visitor evaluation and market research consultancy 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).

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. These gut instincts seem to have a strong optimism bias in my experience, assuming (wrongly) that smiling or interested-looking faces signify effective programme outcomes. 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. In lieu of robust evaluation, the approach presented in the article is essentially an anecdote gathering exercise focused on eliciting positive accounts of how wonderful this programme is. This is not evaluation in any meaningful sense, but rather an advocacy or campaigning activity that belongs in a public relations department. 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 an 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 and 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].' [Humphreys and Jensen, in press]


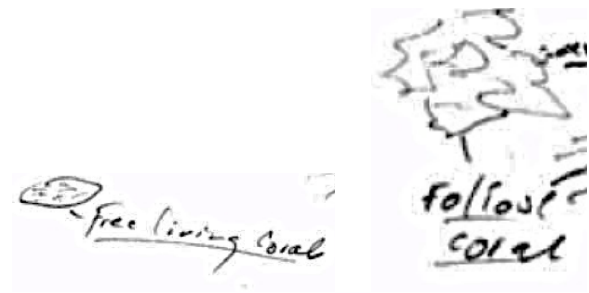
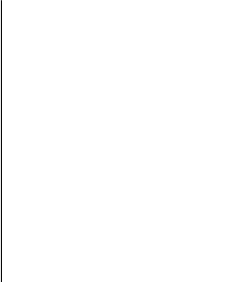
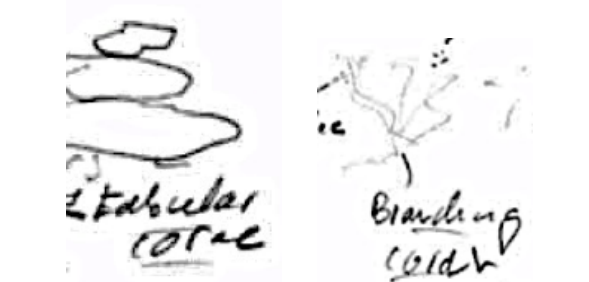
In addition to feedback, I directly measured impact using qualitative methods. There was evidence of transformations in delegates' thinking from pre- to post-training. The main patterns of positive change evident in the thought-listing data (e.g. qualitative responses to 'What comes to mind when you 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 low-budget example of impact evaluation using primarily qualitative methods highlights that direct, robust measurement of intended knowledge outcomes is in fact feasible.

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.

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. Given that the impact of the CPD event on the teachers who attended it was not evaluated, the following hypothetical imagined scenario is fully consistent with the above finding:

A badly designed, poorly conceived CPD intervention at the Science Museum led to teachers misunderstanding key messages about best practice in teaching science and learning new practices that are likely to result in negative impacts on pupils' learning. They have then taken these bad ideas back to the school and told other people about them, encouraging them to adopt the same poor practices in their classrooms, thus spreading the negative impact of the Science Museum 'Talk Science' CPD programme throughout the school.

Given that the evaluation method described in King et al. [2015] cannot tell us whether the above 'impact' scenario 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 affecting pupils' assessment of the teaching and learning activities in their class and their understanding of the topics targeted by the intervention).

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 and 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 (Figure 1). 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].

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:



Figure 1. Pre- and post-visit drawings for same child from impact evaluation at London Zoo [Jensen, 2014a].

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 approach is obviously 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 (including those presented as useful by King et al.) 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 and 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: *The SAGE International Handbook of Educational Evaluation*. Ed. by K. E. Ryan and J. B. Cousins. Thousand Oaks, CA, U.S.A.: SAGE, pp. 97–113.
- Davies, M. and Heath, C. (2013). *Evaluating Evaluation: increasing the Impact of Summative Evaluation in Museums and Galleries*. URL: <http://visitors.org.uk/files/Evaluating>.
- Dawson, E. and Jensen, E. (2011). 'Towards a 'contextual turn' in visitor research: evaluating visitor segmentation and identity-related motivations'. *Visitor Studies* 14 (2), pp. 127–140.
- Humphreys, L. and Jensen, E. (in press). 'Mobilising tourism for conservation education'. In: ed. by E. Jensen. Cambridge, U.K.: Cambridge University Press.
- Jensen, E. (2011). 'Evaluate impact of communication'. *Nature* 469, p. 162.
- (2014a). 'Evaluating children's conservation biology learning at the zoo'. *Conservation Biology* 28 (4), pp. 1004–1011.
- (2014b). 'The problems with science communication evaluation'. *JCOM* 13 (01), C04. URL: http://jcom.sissa.it/archive/13/01/JCOM_1301_2014_C04/.
- Jensen, E., Dawson, E. and Falk, J. (2011). 'Dialogue and synthesis: developing consensus in visitor research methodology'. *Visitor Studies* 14 (2), pp. 158–161.
- Jensen, E. and Holliman, R. (2009). 'Investigating science communication to inform science outreach and public engagement'. In: *Investigating science communication in the information age: Implications for public engagement and popular media*. Ed. by R. Holliman, L. Whitelegg, E. Scanlon, S. Smidt and J. Thomas. U.K.: Oxford University Press, pp. 55–71.
- King, H., Steiner, K., Hobson, M., A., R. and Clipson, H. (2015). 'Highlighting the value of evidence-based evaluation: pushing back on demands for 'impact''. *JCOM* 14 (02), A02. URL: http://jcom.sissa.it/archive/14/02/JCOM_1402_2015_A02.
- Moss, A., Jensen, E. and Gusset, M. (2015). 'Evaluating the contribution of zoos and aquariums to Aichi Biodiversity Target 1'. *Conservation Biology* 29 (2), pp. 537–544.
- Owen, J. and Rogers, P. J. (1999). 'Impact evaluation'. In: *Program evaluation*. Ed. by J. M. Owen and P. J. Rogers. London, U.K.: SAGE, pp. 263–308.

- Trench, B. (2014). 'Do we know the value of what we are doing?' *JCOM* 13 (01), E.
URL: http://jcom.sissa.it/archive/13/01/JCOM_1301_2014_E.
- Wagoner, B. and Jensen, E. (2010). 'Science learning at the zoo: evaluating children's developing understanding of animals and their habitats'. *Psychology & Society* 3 (1), pp. 65–76.
- (2015). 'Microgenetic evaluation: studying learning in motion'. In: *The Yearbook of Idiographic Science. Volume 6: reflexivity and Change in Psychology*. Ed. by G. Marsico, R. Ruggieri and S. Salvatore. Charlotte, NC, U.S.A.: Information Age Publishing.

Author

Dr. Eric Jensen is an Associate Professor of Sociology, University of Warwick. He teaches on social research methods, audience research and science communication. He has conducted numerous impact evaluation studies, including at the Natural History Museum (London), National Gallery (London), World Association of Zoos and Aquariums and the University of Cambridge Museums. His research has been published in journals such as *Nature*, *Public Understanding of Science* and *Conservation Biology*. Jensen has a Ph.D. in sociology from the University of Cambridge. E-mail: E.Jensen@warwick.ac.uk.

How to cite

Jensen, E. (2015). 'Highlighting the value of impact evaluation: enhancing informal science learning and public engagement theory and practice'. *JCOM* 14 (03), Y05.



This article is licensed under the terms of the Creative Commons Attribution - NonCommercial - NoDerivativeWorks 4.0 License.
ISSN 1824 – 2049. Published by SISSA Medialab. <http://jcom.sissa.it/>.