

Highlighting the value of evidence-based evaluation: pushing back on demands for ‘impact’

Heather King, Kate Steiner, Marie Hobson, Amelia Robinson and Hannah Clipson

Abstract

This paper discusses the value and place of evaluation amidst increasing demands for impact. We note that most informal learning institutions *do not* have the funds, staff or expertise to conduct impact assessments requiring, as they do, the implementation of rigorous research methodologies. However, many museums and science centres *do* have the experience and capacity to design and conduct site-specific evaluation protocols that result in valuable and useful insights to inform ongoing and future practice. To illustrate our argument, we discuss the evaluation findings from a museum-led teacher professional development programme, Talk Science.

Keywords

Science centers and museums; Public engagement with science and technology

Introduction

Like all organisations supported by private or governmental funds, museums and science centres are increasingly being asked to document the ‘impact’ of their public service provision. Such calls reflect the current emphasis in government policy discourse and the growing demand for organisations to demonstrate ‘value’ [Kail, Vliet and Baumgartner, 2013]. In this paper, we question the feasibility and usefulness of such demands. Furthermore, we ask whether museums and other informal settings have the skills and capacity to measure ‘impact’. Although we recognize that funders have a right to know to what ends their money is spent, we query what may be realistically documented. In particular, we argue 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. Moreover, we suggest that the emphasis on impact is obfuscating the valuable role of evaluation.

In contrast to impact analyses, evaluation studies focus on immediate questions of what and how something works. Different forms of evaluation — from front-end, through formative and remedial to summative — direct attention to the varying stages of a process, but overall, all forms seek to identify what could be improved, what barriers need to be removed, and what lessons may be taken forward [East of England Museum Hub, 2008; Friedman, 2008].

To illustrate our argument, we discuss the evaluation findings from a museum-led teacher professional development programme, Talk Science, developed by the Science Museum, London and which initially ran for a five-year period from 2007–2012. During this time a number of studies, both quantitative and qualitative, were conducted for both funders and internal stakeholders. In reviewing findings from these studies, we discuss what may be realistically delivered, note the inability to claim ‘impact’, and highlight some of the many valuable insights generated from the evaluation data. In reflecting upon an example of practice in this way, we support calls for greater discussion on the nature of evaluation and documentation of impact [Michalchik and Gallagher, 2010; Pekarik, 2010]. We hope that the discussions developed here will be of interest to theorists, practitioners and policy makers across the field of learning in informal settings.

What is possible to measure?

In recent years, informal learning institutions have been the recipients of substantial sums of money from trusts, non-governmental organisations, and commercial companies, as most major projects require funding additional to core grant-in-aid support. In accepting such funds, recipients are generally contracted to report on how the money was spent, and whether the initiative ‘worked’ and to what extent, i.e. its ‘impact’. Understandably, both sides are keen for the initiative to have been successful, and as a consequence, staff can become trapped by the need to demonstrate impact. Museum consultants have previously identified categories of impact with the aim of helping practitioners identify change. Such categories include changes in visitor awareness, knowledge and understanding, engagement or interest, attitude, behaviour and skills [Friedman, 2008]. Such categories are purposively broad to capture the full range of possible effects. But herein lies the rub: detecting and thereafter measuring change across such a spectrum is no small order.

Some institutions, including the London Science Museum, conduct extensive evaluation studies that guide both the development of initiatives and their subsequent review. Most studies, however, are not designed to demonstrate impact. There are several reasons for this. Firstly, and as noted above, informal learning institutions rarely have the staff, time or budget to conduct the types of research necessary — longitudinal; large-scale; involving control populations; with both quantitative and qualitative elements — to generate appropriate data from which statistically validated claims of impact may be made. On occasion, academic researchers are invited to study initiatives and examine visitor engagement (e.g. UPCLOSE, the University of Pittsburgh Centre for Learning in Out of School Environments <http://uplclose.lrdc.pitt.edu>), but these instances are relatively rare.

Secondly, and more fundamentally, it is very difficult to disaggregate the effect of learning experiences offered by an institution from other related experiences [Anderson, Lucas and Ginns, 2003]. With particular reference to teacher professional development programmes, researchers have argued that many experiences contribute and combine to change teacher knowledge and action and we cannot isolate just one experience [Grenier, 2010]. Moreover, we know that teacher development is complex, potentially involving changes not only in knowledge and specific actions, but also in beliefs and general behaviour [Glackin, 2013]. Proving the causal effect of one particular intervention (over and above other experiences) and thereafter claiming impact is, therefore, extremely difficult.

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. With regards to documenting the success of a museum-led teacher professional development programme, we argue that more qualitative approaches are necessary to firstly begin to understand the nature of developments within individual teachers and then to map these to the training provided. It is only when the unique nature of a professional development exercise are understood — from its design, to its delivery and ultimately its potential effects on teacher practice — that it might be possible to undertake a longitudinal study (perhaps involving comparison of participants of two professional development experiences) and thereafter seek to make a claim of long-term impact.

Lastly, it is important to note that the ultimate objective of teacher professional development is to enhance the learning experience of students. Thus for any claims of impact to be made, changes must be detected in a population (of students) separate from those experiencing the intervention (the teachers). Whilst we acknowledge that some researchers have established a causal relationship between teacher training approaches and student attainment (see Adey, 1993 and Shayer, 1999 for findings from CASE), we note that their research methodology involved longitudinal analyses of a large population of learners, a feat not easy for a lone informal learning institution to achieve.

In sum, there is a tension between what informal learning institutions can feasibly do, and the claims of impact that many are being asked to make. To resolve this tension, institutions need to be clear as to what they can realistically do, and moreover highlight the potential valuable insights that may instead be identified by staff experienced in conducting evaluations. Indeed, we argue that there is a need to push back on unrealistic demands for 'impact' that show demonstrable increases in, for example, audience learning, and instead identify the areas in which it is possible to detect opportunities for improvement in staff practice or exhibit design. By improving practice in this way, better experiences for learners are inevitable. In addition, we argue that it is essential that the field, encompassing funders, practitioners and policy-makers, develop and employ a new discourse which acknowledges the place of evaluation and limits calls for longer-term impact assessment to those that have the capacity and capability to conduct such requirements.

We turn now to a discussion of the evaluation of the Talk Science teacher professional development programme with the aim of exploring elements of this new discourse in practice.

The Talk Science teacher professional development programme

In 2006 BP gave the Science Museum in London just over one million pounds to support a three-year professional development programme for secondary science teachers entitled Talk Science. Further funding of nearly £600,000 was provided in 2009 to extend the project for another two years. The foundational rationale of the programme was to incorporate many of the elements and practices used by the Museum's Contemporary Science Team (for example, in their production of the Antenna Rapid exhibitions which are quickly updated in response to science in the news), and by the Museum's programme of dialogue events for adults. The Museum has developed considerable experience in these areas since the development of the contemporary science Wellcome Wing in 2000, and staff in the Learning team had long wanted to apply this expertise to support learners beyond the Museum's walls. Thus the overall aim of the Talk Science professional development programme was to develop ways of supporting teachers to increase student engagement with contemporary science. It aimed to provide teachers with the tools and techniques to run contemporary science discussions in the classroom to help give students the confidence to find their own voice and have a say in the way science impacts and shapes their lives.

The Talk Science programme comprised a one-day training session for teachers and addressed facilitation techniques, discussion formats and classroom activities, and included opportunities for teachers to share ideas and best practice. The day was supported by a website and blog that was updated for access by teachers after the course. Talk Science also included the development of new resources for teachers, plus a suite of online games to support dialogue activities with a teenage audience. Over the five years, 152 Talk Science courses were run in 30 cities and towns in the UK. In total, 2625 secondary science teachers participated in some aspect of the Talk Science Professional development programme.

Evaluating Talk Science

The evaluation of Talk Science comprised three stages: front-end, formative and summative. Each stage of evaluation, together with the specific tasks undertaken, is described below.

4.1 Conducting 'front end' evaluation to understand and identify existing best practice

Front end evaluation may be defined as evaluation in the planning stages of an exhibition or programme to assess the knowledge, misconceptions, interests, attitudes and preferences of potential visitors [Bitgood and Shettel, 1996]. It can also help to identify knowledge about existing research and good practice, in the wider field and in related areas, upon which to draw.

The key front end task conducted for Talk Science was an initial literature review to identify a possible theoretical perspective for framing the professional development efforts. The Talk Science team selected the framework developed by Thomas Guskey [2000] as the most appropriate for their needs. Guskey's framework describes five incremental levels of outcome for a 'successful' teacher professional development experience thus:

1. Participating teachers have a positive learning experience.
2. Participating teachers gain new knowledge and skills.

3. Change is supported in the participating teacher's school.
4. The new teaching approaches are implemented.
5. Gains are visible in student learning.

4.2 Formative evaluation — identifying and removing barriers to learning and potential impact

Formative evaluation is used to provide additional visitor input into the development process. Inexpensive versions or prototypes are tested with visitors and revised and retested as necessary until they effectively communicate their intended messages to their intended audience [Bitgood and Shettel, 1996].

To conduct the formative evaluation, a dedicated Audience Researcher was recruited. The researcher developed a methodology and tools for the formative tests comprising a mixture of quantitative feedback surveys and in-depth telephone interviews. Other formative evaluation studies included testing teacher reactions to the resource materials designed to accompany the course [see Hobson, 2008]. The formative evaluation studies identified barriers to learning and made recommendations for amendments.

4.3 Summative evaluation: end of programme reporting

Summative evaluation follows the completion of a project and assesses the extent to which the project has met its intended outcomes for its intended audience. It can include recommendations for remedial actions for the future and should include broader learning points for future projects.

The summative evaluation for Talk Science was based on Guskey's model of success identified in the front-end evaluation stage. A semi-structured questionnaire was developed involving a number of items that mapped onto Guskey's five levels of success as indicated in the table below.

The questionnaire was sent to 980 teachers who had attended a Talk Science course between 2007–2012. 79 responses were received. This is acknowledged to be an extremely low return rate. Moreover, it was not possible to determine how representative of the wider set these 79 were: they may have been the most enthusiastic participants, and thus their responses may be positively skewed. However, the quantitative data was supplemented by six in-depth telephone interviews and three face-to-face interviews with teachers, each lasting approximately 45 minutes. This helped to provide a more rounded picture of the respondents' views. All data were collected in July 2012.

Analysis of findings

As noted earlier, a low number of responses limits the ability to conduct any statistical analyses. In addition, there were no data from a control group, or collected from the group prior to their participation in the programme. Moreover, whilst Guskey's five stages provided staff with a useful framework for conceptualizing the professional development process, collecting the data to

Table 1. The question areas used for the evaluation of the Talk Science course mapped against Guskey’s five levels of outcomes for an effective professional development programme.

Question areas	Guskey’s Levels of Outcomes for successful professional development (2000)
What are teachers’ levels of satisfaction with the course? Did the course meet their needs? Would the teachers take part in further Talk Science courses?	Level 1: Participants have a positive learning experience.
What did teachers report learning by attending the course with particular regards to: <ul style="list-style-type: none"> – Teaching discussion. – Teaching contemporary science. – The Science Museum as a resource for teachers? Do they feel more confident and competent to teach contemporary science topics and to use discussion techniques in their teaching?	Level 2: Participants gain new knowledge and skills.
How have teachers disseminated Talk Science tools and techniques within their schools? How supportive are teachers’ schools in enabling them to implement Talk Science tools and techniques?	Level 3: Change is supported in the participating teacher’s school.
How have teachers implemented the Talk Science tools and techniques in the classroom?	Level 4: The new teaching approaches are implemented.
What do teachers say about their students’ response to the Talk Science tools and techniques? What do the teachers say about how student attitudes changed as result of the new tools and techniques being implemented in class? Have teachers seen any improvements in other aspects of their students’ behaviour, actions or learning as a result of these techniques?	Level 5: Gains are visible in student learning.

confidently claim changes in for, example, teacher practice and thereafter student learning (levels 4 and 5) was acknowledged to be impossible. However, the Guskey levels provided the framework against which teacher self-reports could be matched, and thus provided the team with a source of insight and suggestions for improvement as outlined below.

Meeting Guskey’s level 1: Participating teachers have a positive learning experience

From the analysis of their responses the sampled teachers were clearly satisfied with the programme: 98% agreed or strongly agreed that they were inspired; and 98% found the course to be relevant to their work. On probing which element of the programme was most highly valued, four main themes could be detected:

- The engaging and interactive nature of the course
- The opportunity to network and share ideas with fellow teachers
- The ability to immediately implement the new resources, techniques and ideas
- The practical relevance of the course for day-to-day teaching

These programme-specific findings are reassuring for the designers for the Science Museum’s Talk Science course and indeed were useful in informing decisions about revisions to the course structure. Moreover, even despite the low sample size, it is also possible to interpret these evaluation results more generally. For example, the findings underscore the potential strengths of informal learning settings for delivering professional development to teachers. Good presentation skills and an interactive approach are hallmarks of informal learning settings and in this instance Science Museum staff used a number of contemporary science formats developed since the opening of the Wellcome Wing and Dana Centre including, for example, controversial drama characters such as the pregnant man, in order to engage their audiences. A possible conclusion here is that when novel resources are combined with facilitation techniques typical of informal learning settings the potential for creating a satisfying learning experience is high. This conclusion is of relevance to the field and is an example of how evaluation studies conducted in one setting can be applied elsewhere.

Meeting Guskey’s level 2: Participating teachers gain new knowledge and skills; and Gusky’s level 4: The new teaching approaches are implemented

Meeting Guskey’s second level requires the acquisition of new knowledge and skills. If any such acquisition was indeed found, a claim of impact could conceivably be made. However, as discussed above, detecting an increase in knowledge is not easy and in this case was not possible. For example, the teachers did not take a pre-test of their knowledge prior to participating in the course which could be compared with a post-test assessment. Moreover, whilst teachers in both the questionnaires and interviews noted that they had learnt new techniques and had applied them to address particular needs, the limitations of self-report data must not be forgotten [Donaldson and Grant-Vallone, 2002]. If, on the other hand, we view the feedback from an evaluation perspective, the responses offer useful insights. For example, and in reference to level 4, several teachers described ways in which they had adapted the Talk Science techniques across a broad range of topics including ethics, reproduction, particle theory, vaccination and stem cells. Their approaches have since been shared more widely by the Science Museum, and indeed have informed the development of related courses.

Meeting Guskey level 3: Change is supported in the participating teacher’s school

90% of respondents commented that they had actively disseminated new knowledge and skills to their colleagues back at school. This was achieved either through departmental meetings, in advising student teachers or through informal discussions — feeding back in an ad hoc way, giving tips on good ideas and so on.

However, whilst many teachers reported sharing their newly acquired insights, most admitted that the Talk Science techniques have not been fully taken up by colleagues and that organizational change has not occurred. This finding concurs with prior research that indicates that the products of professional development — new ways of thinking and teaching — take time to become embedded by the participant teacher, and longer to be shared with and taken up by other colleagues [Loucks-Horsley et al., 2002].

As a result of identifying this barrier to dissemination as part of the formative evaluation, the Science Museum team explored the potential of developing online presentations for participants to use to communicate the experiences of the course to colleagues. In this instance, therefore, an evaluation approach used findings to identify and overcome barriers to learning for future iterations of the course. An 'impact' approach, on the other hand, would have sought to measure change in organizational practice as a direct result of the course, and if non-existent a conclusion of failed delivery would be likely.

Meeting Guskey's level 5: Gains are visible in student learning

Finding evidence of Guskey's level 5 is, for many of the reasons discussed already, particularly hard to demonstrate. The evaluation conducted did not include any pre and post observations of teaching practice or any measures of student attainment that would be required in order to triangulate the teachers' perceptions of gain. 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

From the findings presented above, it is clear that the teachers valued their experience of participating in the Talk Science programme. Furthermore, they believed that the programme was of benefit to their students. From the Science Museum's perspective, this is, of course, very positive. It is also very positive for the programme's funders: while long-term impact may not have been definitively proved, the evaluation indicates money and time well spent. The Talk Science evaluation also served to identify a number of potential barriers that may prevent learning and the subsequent implementation of practice, such as the length or timing of a professional development course. For example, 83% respondents commented that they would like to attend the same course or similar courses again, but many felt that it would be difficult to leave school for a whole day. The implications of this finding have already been explored with evening courses being considered for the future.

Final words

In this paper, we have highlighted the difficulty faced by informal learning institutions in claiming impact given their limited resources and constraints (financial and otherwise) for conducting the necessary studies for generating the appropriate data. However, we have also demonstrated that evaluation studies — such as that conducted for Talk Science — can offer important insights for the institutions concerned.

We are thus calling for the informal science sector to push back on unrealistic calls from funders and government agencies to identify long-term impact. Instead, the sector should emphasise the value that may be afforded by evaluation studies, and thereafter work to ensure that evaluation aims and approaches are realistic. Evaluation efforts of this sort will deliver useful and tangible findings in the form of site-specific insights but could also serve to highlight issues and considerations for the wider field.

In discussing the constraints faced by informal learning institutions, we would also like to draw attention to the need to develop greater research capacity in the sector in order to ensure that evaluations are conducted in the best ways possible: that theory is used to frame evaluation (in the way that the Science Museum applied Guskey's five steps in order to benchmark their efforts); and that the guiding questions are developed appropriately and in accordance with data collection tools available. In addition, there may be a need to provide staff with further skills in how to conduct evaluation and garner the types and volumes of data necessary for statistical analysis. Indeed, this work has begun — see for example the Ecsite special interest group on evaluation dedicated to supporting informal practitioners develop and utilise appropriate evaluation approaches (http://www.ecsite.eu/activities_and_resources/thematic_groups/rev-group); or the guidance and resources provided by the Center for Advancement of Informal Science (CAISE) (<http://informalscience.org/evaluation>).

Furthermore, we note initiatives currently underway promoting closer relationships between academic researchers and informal sector evaluators [King and DeWitt, 2013]. In short, it would appear that the informal science sector is becoming increasingly well placed and well equipped to redefine expectations from funders. A new tone for what constitutes reasonable and feasible evaluation is emerging. It is our hope that this new tone spreads across the sector and informs the efforts and expectations of both institutions and their funders.

A footnote

Since this paper was first formulated, BP announced the funding of a five-year programme at the Science Museum which included the development of further professional development opportunities which began in the summer of 2013. The funding also supports a research partner in the form of King's College London to both advise on the development of initiatives and to design a rigorous research programme which will combine the numbers, approaches, and analyses necessary to claim impact. Such research is being conducted alongside an in-house evaluation of the programme, the latter providing the valuable insights that are so important for ongoing practice.

References

- Adey, P.S. (1993). 'Accelerating the development of formal thinking in Middle and High school students IV: three years on after a two-year intervention'. *Journal of Research in Science Teaching* 30 (4), pp. 351–366.
- Anderson, D., Lucas, K.B. and Ginns, I.S. (2003). 'Theoretical perspectives on learning in an informal setting'. *Journal of Research in Science Teaching* 40 (2), pp. 177–199.
- Bitgood, S. and Shettel, H.H. (1996). 'An overview of visitor studies'. *Journal of Museum Education* 21 (3), pp. 6–10.
- Donaldson, S.I. and Grant-Vallone, E.J. (2002). 'Understanding self-report bias in organizational behaviour research'. *Journal of Business and Psychology* 17 (2), pp. 245–260.
- East of England Museum Hub (2008). Evaluation toolkit for museum practitioners. Norwich, U.K.: East of England Museums Hub.
- Friedman, A. (2008). Framework for evaluating impacts of informal science education projects. URL: http://informalscience.org/documents/Eval_Framework.pdf.
- Glackin, M. (2013). 'Teaching science outside the classroom: the role of teachers' beliefs and teacher efficacy during a two-year professional development programme'. Ph.D. thesis. London, U.K.: King's College.
- Grenier, R.S. (2010). "'Now this is what I call learning!'" A case study of museum-initiated professional development for teachers'. *Adult Education Quarterly* 60 (5), pp. 499–516.
- Guskey, T.R. (2000). Evaluating professional development. Thousand Oaks, CA, U.S.A.: Corwin Press.
- Hobson, M. (2008). Teachers reactions to the Talk Science Products: concepts and first prototypes. Unpublished report. London, U.K.: Science Museum.
- Kail, A., Vliet, A. Van and Baumgartner, L. (2013). Funding Impact. Impact measurement practices among funders in the U.K. London, U.K.: NPC.
- King, H. and DeWitt, J. (2013). A framing narrative for collaborative partnerships supporting research and learning in informal science institutions. CPPR Working Papers Series. Paper No. 6. URL: <http://www.kcl.ac.uk/sspp/departments/education/research/cppr/workingpapers/index.aspx>.
- Loucks-Horsley, S., Hewson, P., Love, N., Stiles, K. and Mundry, S. (2002). Designing professional development for teachers of science and mathematics. 2nd ed. Thousand Oaks, CA, U.S.A.: Corwin Press.
- Michalchik, V. and Gallagher, L. (2010). 'Naturalizing Assessment'. *Curator* 53 (2), pp. 209–219.
- Pekarik, A.J. (2010). 'From knowing to not knowing: moving beyond "outcomes"'. *Curator* 53 (1), pp. 105–115.
- Shayer, M. (1999). 'Cognitive acceleration through science education II: its effects and scope'. *International Journal of Science Education* 21 (8), pp. 883–902.

Authors

Heather King is a Research Fellow at King's College London. Her research focuses on the role and function of informal science learning institutions and, in particular, the pedagogical practices of informal educators. She also has extensive experience of designing and conducting evaluation studies for a number of EU funded initiatives. E-mail: heather.1.king@kcl.ac.uk.

Kate Steiner is currently Editor of the new Science Museum Group Journal, an open-access online journal publishing scholarly articles relating to the work and

interests of the Science Museum Group. Kate was previously head of Audience Research at the Science Museum Group and has led many research and policy initiatives in the field of visitor studies.

E-mail: Kate.Steiner@sciencemuseum.ac.uk.

Marie Hobson is Learning Researcher at the Natural History Museum, London, and Chair of the Research and Evaluation Thematic Group, Ecsite. She was an Audience Researcher on the Talk Science project. E-mail: M.Hobson@nhm.ac.uk.

Amelia Robinson is Research Manager at the Heritage Lottery Fund. She was an Audience Researcher on the Talk Science project.

E-mail: Amelia.Robinson@hlf.org.uk.

Hannah Clipson is Community Learning Officer at The British Postal Museum & Archive. She was an Audience Researcher on the Talk Science project.

E-mail: Hannah.Clipson@postalheritage.org.uk.

How to cite

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), A02.



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/>.