

**RESEARCH FOR ALL**

The open-access journal for public engagement with research

ISSN 2399-8121 (Online)

Journal homepage:

<https://www.uclpress.co.uk/pages/research-for-all>

Evaluating the short-term and long-term impact of an interactive science show

Wendy J. Sadler 

How to cite this article

Sadler, W.J. (2021) 'Evaluating the short-term and long-term impact of an interactive science show'. *Research for All*, 5 (2), 399–419. <https://doi.org/10.14324/RFA.05.2.14>

Submission date: 29 August 2020

Acceptance date: 28 February 2021

Publication date: 21 September 2021

Peer review

This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymized during review.

Copyright

© 2021 Sadler. This is an open-access article distributed under the terms of the Creative Commons Attribution Licence (CC BY) 4.0 <https://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Open access

Research for All is a peer-reviewed open-access journal.



Evaluating the short-term and long-term impact of an interactive science show

Wendy J. Sadler* – *Cardiff University, UK*

Abstract

Science shows as a medium for communicating science are used widely across the UK, yet there is little literature about the long-term impact they may have. This longitudinal study looks at the short-term and long-term impact of the science show *Music to Your Ears*, which was initially performed throughout the UK on behalf of the Institute of Physics in 2002, and which has since been offered at schools and events through the enterprise Science Made Simple. The impact was measured using the immediate reaction to the show, the number (and type) of demonstrations (demos) recalled over the long term, and the applied use of any memories from the show. Quantitative and qualitative data were gathered using questionnaires immediately after the show and focus groups held two and a half years later. To enrich the data, and minimize bias, interviews with professional science presenters were also included in the data analysis. Data from the questionnaires were used to develop a framework of five demonstration categories to describe their essence, or main purpose. The categories used in this study were: curiosity (C), human (H), analogy (A), mechanics (M) and phenomena (P). It was found that even after two and a half years, almost 25 per cent of demos from the show could be recalled without prompting. When prompted with verbal and visual clues, over 50 per cent of the demos from the show could be recalled by the group tested. In addition, around 9 per cent of the demos were recalled and related to an alternative context to the show, suggesting that some cognitive processing may have happened with the most memorable elements of the show. The 'curiosity' type of demo was found to be the most memorable in both the short term and long term.

Keywords: science demonstrations, science performance, long-term impact

Key messages

- Demonstrations in a science show can be memorable over the long term – especially those that are counter-intuitive, or curious in nature.
- By categorizing the type of demonstration and comparing short-term and long-term memories, we can predict which types of demonstration may suit different types of learner.
- There is evidence that some short-term memories from demonstrations are applied in other contexts over the long term, which suggests that some cognitive processing of the memory has occurred.

Science shows as a format for public engagement

The history and development of the science show

The demonstration lecture has been an informal method of science communication for many years (O'Brien, 1991). Since around 1700, the public have been attending lectures on 'natural philosophy' and 'valuing science as a momentous part of high culture' (Knight, 2002: 217). Michael Faraday presented science lectures for the public from the 1820s at the Royal Institution, and firmly believed that they were an important way to engage wider audiences and children with science (James, 2002). Most importantly, Faraday recognized that using live demonstrations was an essential part of communicating science. He instigated the Friday Evening Discourses and the Christmas Lectures for children, which still continue today. More recently, since the 1980s, there has been a steady growth in the number of science centres operating in the UK and worldwide. Many centres began to realize that after the extensive spend on the exhibitions, they needed to provide something else to enrich the visit and to bring back repeat visitors. The science show is therefore the modern interpretation of the traditional lecture demonstration. It has evolved over the last thirty years or so into a range of different styles and formats. It is a genre that you can find at a museum and a shopping centre, and many places in between, and it is no longer always presented by scientists. For the purpose of this study, we shall be referring to 'science demonstration lectures' as 'science shows'.

A science show can go to audiences and places to try to reach those 'publics' who may never choose to visit a science centre or museum. For this reason, science shows are an integral part of the movement to increase public engagement with science and technology, and a useful tool to enrich formal science teaching.

How do we define and measure the 'impact' of science shows?

'Impact' is defined in the dictionary as 'a forceful consequence, a strong effect', or something that 'influences strongly'. In the context of this research, we are interested in any 'strong effect' that the show may have had on the people who saw it. The impact was assessed in the short term and the long term on the assumption that truly successful impact creates a positive long-term memory. Opinions were sought immediately after the event, and then from a smaller number of people two and a half years later. A comparison will be made between the things that they enjoyed immediately after the event, and the things that they could recall at a later time. On this occasion, no pre-event measurements were taken, so we cannot compare audiences' initial attitudes or knowledge with the short-term and long-term impact recorded.

What is the show that is being evaluated?

In this study, the science show being evaluated was written and presented by the author on behalf of the Institute of Physics National Schools Lecture Tour (2001/2). (It has since been offered at schools and events through the enterprise Science Made Simple.) The show was called *Music to Your Ears*, with the strapline, 'The story of sound from synths to CDs'. The show is a topic-linked demonstration ('demo') lecture on the subject of sound, waves, music and music technology. It is 60 minutes long, and was originally written for Key Stage 3 and Key Stage 4 students (11–16 years old). The content and show structure are outlined in Table 1.

Table 1: Summary of show content for *Music to Your Ears*

Section	Content covered	Demonstrations
A: What is sound? Duration: 10 minutes	<ul style="list-style-type: none"> – Sound is made when something vibrates – Sound travels as a longitudinal wave – Sound needs a medium in which to travel – Humans can only hear sounds from around 20–20,000 Hz 	<ul style="list-style-type: none"> – Feeling vibrations of voice box – Longitudinal wave on a slinky spring – Audience making vibration with hands
B: How we hear? Duration: 10 minutes	<ul style="list-style-type: none"> – Process of sound travelling into ear via eardrum, hammer, anvil, stirrup and then into cochlea – How hairs in the inner ear transfer vibrations into electrical signals to the brain – What tinnitus is, and how it happens – How animals have different hearing ranges to us 	<ul style="list-style-type: none"> – Audience make larger ears using hands to hear the difference in sound quality – Audio demo of tinnitus – Hearing test to find out highest sound the audience can hear
C: What is music? Duration: 20 mins	<ul style="list-style-type: none"> – That you need an ordered, periodic waveform to create something that sounds like music – High-pitched sounds have more vibrations or waves per second than low-pitched sounds – Loud sounds have large amplitude waves, quiet sounds have small amplitude waves – Musicians call things volume and pitch; scientists call them amplitude and frequency. One is subjective the other objective – That 'noise' has a mixture of lots of frequencies and does not really sound musical – That you need resonance to maintain a musical sound – That you can use friction to make something vibrate (wine glass or violin) – That turbulent air can set up a resonant vibration – Things vibrate, or resonate at speeds that relate to their length, or mass – Large things usually resonate at low frequency so we get low notes, small things resonate at higher frequency to give us higher notes – Instruments have more than one possible resonant frequency, which are called harmonics 	<ul style="list-style-type: none"> – Volunteer sings into digital oscilloscope to see shape of wave live on screen – Volunteer shows noise, and different pitches on oscilloscope – Make wine glass resonate by rubbing with wet finger – Use blowtorch to excite resonance in metal pipe – Show sticks of different lengths resonating at different speeds – Musical drinking straw making vibration that changes with the length, cutting off pieces as you blow to get rising pitch – Whirly tube, spun around head to show different harmonics of tube – Standing waves on string driven by signal generator, illuminated by strobe light to freeze-frame motion

Table 1: (continued)

Section	Content covered	Demonstrations
D: Technology and music Duration: 15 minutes	<ul style="list-style-type: none"> – That harmonics are related mathematically; if lowest resonate frequency is F_0, then the next one can be found at twice that frequency, $2F_0$, and so on – That all sounds have unique fingerprints of harmonics that your ear and brain detect to tell you what is making the sound – That generally you can amplify a sound acoustically by making more air vibrate – using a sound board – That you can use electronics to amplify a sound – That you can use electronics to change the pitch of a sound – That you can generate musical sounds using electronics, and copy acoustic instruments using synthesizers – That analogue recording is an exact copy of the original acoustic signal and is a continuously varying signal (how records work) – That digital music is a way of transferring continuously varying signals into 0s and 1s (or ons and offs), which can be done by sampling (how CDs store music) – That MP3 files use the weaknesses of the human ear to get rid of excess information or sounds that we cannot hear very well, which is how they can be compressed to much smaller data sizes – Technology has not been too successful at copying a human-sounding voice yet, but voice synthesis can be a way of getting computers to speak, or even sing to you. 	<ul style="list-style-type: none"> – Play tiny music box and show how sound gets much louder when placed on a table or tray – Audio demo of mic and amp – Volunteers use voice changer to change pitch of their voices – Show theremin, which is one of the first electronic instruments ever made, played by moving your hands around in the air – Listen to early voice synthesis to see if you can work out words being said – Listen to more recent programs that allows computer to sing melodies according to lyrics that you put in

What does the research tell us?

The development and production of science shows similar to the one examined here is a specialist field driven by a small group of practitioners. Because of this, there is very little academic research that applies directly to the evaluation of science shows in an informal education environment.

One key study that does address the evaluation of science shows as a tool for science communication looked at 36 case studies of science shows being performed in the UK, Australia and the USA, and examined the many different types of show in detail. It looked particularly at the process of writing and evaluating a show, and found that 70 per cent of the case studies chose to use demonstrations and experiments as the primary medium for the shows (Burns, 2003). In addition, the study found that of 110 audience members who had viewed a science show in a shopping centre, the average number of demonstrations that could be recalled (after three weeks) was three. There were 12 demonstrations in total in this instance, and two people questioned could remember 9 of the 12 presented (Burns, 2003). Unfortunately, no detail was given of which demonstrations were remembered on that occasion, so I am unable to use this as a comparison tool to the results of this study. Another useful reference (Bultitude and Eigenbrot, 2004) was a short-term evaluation of a science show called *Lasers Light up Your Life*. This study found that after the show, over 70 per cent of the audience had learned either 'some' or 'a lot' of science. The affective gains were also measured, and it was discovered that the number of audience members who reported 'really liking' science increased by almost 10 per cent as a result of watching the show. These examples provide some evidence to support the anecdotal belief among professionals that science shows are a motivational experience that can have both cognitive and affective gains – at least in the short term.

What do we know about short-term and long-term memory in informal learning?

There have been many studies into the perception of science immediately following a public engagement event or campaign, but what about studies of long-term memory? Some studies have claimed that assessing 'learning' immediately after a visit or event is not realistic, because 'much of what we can be said to learn from an event where new content is presented is dependent on subsequent reinforcing experiences related to that content' (McManus, 1993: 367). Some have tried to address whether any positive impact is sustainable in the long term. In 1991, a research study interviewed visitors to the Launch Pad hands-on gallery in the Science Museum, London. Learning outcomes were assessed immediately after the visit, a few weeks later, and several months later. It was found that the social interaction and subsequent conversations between those present at the exhibit helped make the content of exhibits memorable even up to six months later (Stevenson, 1991). A more recent study of primary school students, where recall of a science centre visit was prompted by photographs and videos, also found a high level of similarity between short-term and long-term exhibit memories (De Witt and Osborne, 2010).

With regard to memory in more general terms, it is believed that there are a number of psychological modes that affect the formation and use of memory, which can be divided into two categories – active modes and passive modes (Herrmann and Plude, 1994).

Active modes – where some action is taken to try to prompt a memory:

- **Mental manipulations:** thought processes that foster encoding or the cueing of retrieval
- **Physical environment:** the perception of aspects of the environment, or the physical use of the environment to facilitate such perception, so as to foster encoding or the cueing of retrieval

- **Social environment:** the perception of aspects of the social environment, or engaging in behaviours that facilitate such perception, so as to foster encoding or the cueing of retrieval.

Passive modes – where no action is taken and memories appear spontaneously:

- **Physiological states**
 - Physical condition: factors of daily living that fatigue a person
 - Chemical state: chemicals consumed (stimulants, medicines)
 - Health: minor and major diseases
- **Emotive states**
 - Attitudinal state: disposition to process certain information, or to interact with certain people
 - Emotional state: feelings of varied intensity, quality and value (positive, neutral or negative)
 - Motivational state: awareness of rewards and punishments for performance of memory tasks.

All of these factors affect the ability to recall memories, so they must be taken into consideration when drawing conclusions about the long-term memory of the show, or indeed the ability of the focus group members to recall specific things about the show. By being able to compare immediate feedback with longer-term feedback, it is hoped that we can discover which specific kind of demonstration leads to a high level of retention. In addition, it will be interesting to see if any social interaction during the show (such as the use of volunteers) leads to a better retention of the event. The definitions of different modes that affect the formation of memories may help to make connections between the most cited demonstration and the type of memory that this is likely to involve.

The use of demonstrations in formal science education

Research has shown that access to a higher frequency of demonstrations and hands-on learning can improve the academic achievement and motivation levels of science students (Stohr-Hunt, 1996). Demonstrations can help students to conceptualize phenomena and relate things to the real world. In a study by Di Stefano (1996), less than 10 per cent mentioned the amusement value of demonstrations, whereas just over a third commented that the demonstrations assisted learning. A further third either described a demonstration in some detail or described what they had learned from it (Di Stefano, 1996).

In addition, it has been shown that using multimedia or computer models can help students provide visual models for phenomena that are usually invisible (Wu et al., 2001). Wu et al. (2001) also found that students who discussed concepts socially and managed to link them to other contexts had more success at retrieving the knowledge at a later date. The focus group data collected will be used to explore the social aspect of this particular science show (that is, what discussions happened after the show) and the ability for audience members to make links with other experiences.

Definition of ‘demonstration’ used in this study

For the sake of clarity, it should be noted that in this context my definition of ‘demonstrations’ included things that may not be thought of as science demonstrations

specifically. For example, the show includes a performance of the theremin (electronic music instrument), and this is not a science demonstration per se. Anything that involved the audience doing an experiment, and where volunteers became part of the demonstration, were also categorized as demonstrations, even though they may not include science equipment.

Methods

What research tools will be used and why are they suitable for this type of project?

The two main tools of social science research are positivistic and phenomenological (Hussey and Hussey, 1997). Positivistic research tends to be objective and deals in mainly quantitative methods that assume people behave in a way that can be reproduced to obtain the same results over and over again. This research has its roots in the natural and physical sciences. Phenomenological research accepts that people are affected in some way by being involved in the research, and that the researcher cannot separate their own views and beliefs when conducting and even designing the research. Despite having a background in the physical sciences, my research method for this project is phenomenological because of the close involvement and personal interest in a show that was both written and presented by the researcher. It is important to acknowledge that this personal knowledge and experience could bias the study, but this was minimized by taking steps to script the focus groups and interviews, and not to join in or steer conversations as they developed.

This project involves a longitudinal study. The initial sample group completed questionnaires in December 2001. A small number of these participants were then gathered for a focus group discussion in July 2004. A true longitudinal study does identical sampling and data collection at various timescales. This project did not repeat the same evaluation after a period of time had passed, but instead developed a follow-up method that was used primarily to assess memory and to help develop some of the hypotheses that were generated from the results of the initial sample study.

To further minimize bias, another element of the study involved data about the awareness that professional presenters have of what demonstrations they choose, and how the audience reacts to them. This was done to help establish whether the categories being chosen would have common ground with those that other professionals use.

Questionnaire design and method

The questionnaires were intended to give a large sample of responses from the audience immediately after seeing the presentation. The sample size was fairly large ($n=171$), and some of the data collected were numeric and suitable for basic statistical analysis. The questionnaires also attempted to use the grounded theory method of research, where there was no hypothesis in mind initially to be tested, and it was hoped that the responses from the open-ended questions would help lead the research and define the direction of the focus groups. In total, 72 questionnaires were filled in by the original group of school students who saw the first show in December 2001. The school involved was a single-sex independent school, which it is noted may not be wholly typical of other state schools. To supplement the data, and to compensate somewhat for this, a further 69 questionnaires were collected on tours of South Africa and North Yorkshire using the same show content and with comparable audience ages. The South African data were only from schools with a similar socio-economic status to the

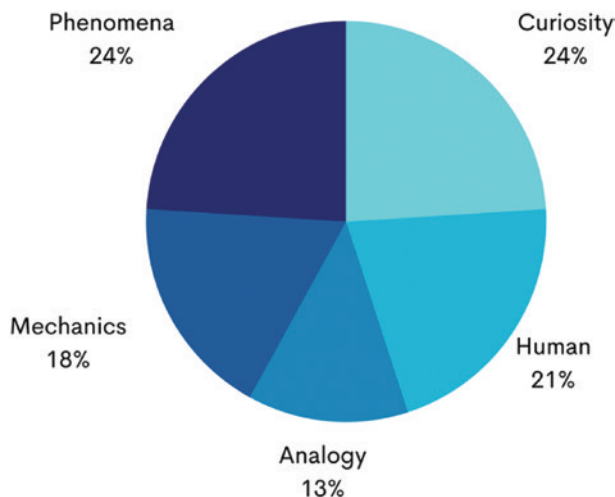
ones in North Yorkshire (rather than from the township schools which were also part of the tour). It should be noted that there could be some cultural differences in science education between students in the UK and South Africa.

Preliminary content analysis was used on the responses to the questionnaires to cross-reference with the data from the focus groups to verify which elements of the show were likely to have the most impact. At this stage, it was important to establish that the spread of demonstration category types (see Table 2) in the whole show was noted, so that if there was a bias of one type of demonstration over another, this could be accounted for in the analysis. The spread of demonstration categories when viewed across the whole show is summarized in Figure 1. The proportion of each

Table 2: Demonstration categories by characteristic

Category code	Category name	Definition	The audience ...?
C	Curiosity	Something weird, a piece of equipment never seen before, something counter-intuitive, a challenge to the audience	Are surprised
H	Human	A volunteer is used, all of the audience take part in an experiment, something personal is learned, something funny happens to the volunteer	Interacts
A	Analogy	A visual representation of something that is usually invisible, using body language to draw a mental picture, using models	Understands
M	Mechanics	How things work, taking things apart and seeing what is inside, how the science is applied in real-life things, how to make simple things, in general applied science – technology	Contextualizes
P	Phenomena	A chance to see a scientific phenomenon happening live, may use equipment not readily available, not necessarily weird, illustrates the basic science using a demonstration rather than words	Experiences

Figure 1: Summary of percentage of demonstration types used in the show (credit: author)



type was roughly comparable, apart from a slightly lower occurrence of 'analogy'-type demonstrations.

To construct a core group of demonstration types, data from the interviews with other professional presenters were used in conjunction with analysis of the questionnaire data. Many demonstrations have dual (or even triple) purpose, but usually the main purpose of the demonstration is clear when considered within the narrative of the show. The five categories established were 'curiosity', 'human', 'analogy', 'mechanics' and 'phenomena'. These categories are not mutually exclusive, and many demos have elements of two or three of the categories; however, there is usually one of the categories that becomes the primary one for each demo. The defining characteristics of each of these categories are given in Table 2.

These categories formed the basis of the content analysis that was implemented on the focus group data. Once the categories were established, the short-term impact from the questionnaires could be analysed. This could then be compared with the focus group data to see if the short-term and long-term impact were related.

Focus groups

The focus groups were designed to follow up the trends that came out of the questionnaire analysis, and therefore they were not scripted until analysis of the data from the questionnaires was complete. In effect, the results of the questionnaires defined the proposed script for the focus groups, so that the results from each could be used together.

There was a potential sample size of 72 questionnaires from the school in Cardiff who saw the preview show. It was decided that a realistic sample size for the focus groups would be around 10 per cent of this. The school selected ten students to take part. Five of these were studying A-level physics; the other five were not.

Both focus groups were transcribed in full, including all the words used in the introduction and by the interviewer. Using the hypothesis that demonstrations fit into certain category types, the transcriptions were coded. The coding system used letters relating to the demonstration category, with additional codes to take note of other memory recalls that were not specifically about a demonstration. These codings of 'other memory' are shown in Table 3.

Initial analysis had shown that some memories showed evidence of applying the content to new contexts, and it was felt that it would be particularly useful to take note of these 'related' mentions, as this is recognized as a measure of 'impact' and memory. In addition, it was necessary to record any 'wrong' memories, because if students remembered a large proportion of things incorrectly, this could be a sign that the show was not constructed well, or was not having the desired impact. Irrelevant memories and memories about the style of the show in general were also included, so that the category memories could be expressed as a percentage of total memories recalled.

Table 3: Explanations and abbreviations for focus group data

Demo category memory	Other memory
C – Curiosity	S – a memory about the style or format of the show, not a demo or part of the show itself
H – Human	R – related memory where the person has made a connection from something in the show to another experience since the show
A – Analogy	X – a memory that does not relate to the show
M – Mechanics	W – a memory about the show that is 'wrong'
P – Phenomena	

The total number of demonstrations mentioned by each group was analysed before prompting, after a verbal prompt, after a pictorial prompt and after a prompt using a prop/real object from the show. This gave a general idea of how many demonstrations in total could be recalled from the show at different times of the discussion.

Interviews with professionals

The British Interactive Event is an annual conference for anyone working in the field of interactive communication. During this conference, semi-structured interviews were conducted with a total of six professional presenters.

The presenters remain anonymous within the study. The aim of doing this extra data collection was to try to establish how the process of demonstration categorization fitted within other professional opinions, and also to enrich the data from my small study with some wider comments about the field of science shows in general.

Results of the evaluation

Data from the questionnaires

A total of 171 questionnaires were collected immediately after the show on three different occasions. The first question was a general rating question ('How would you rate the presentation you saw today?'), with four options (excellent, good, adequate and poor); 100 per cent of the audience responded that the show was either 'excellent' or 'good'. Questions 2 and 3 asked about how entertaining and educational (respectively) they felt the show was on a scale of 1 (low) to 5 (high). The average entertainment score was 4.16 for the Cardiff group, and 4.24 for the South Africa and Yorkshire group. The average educational scores were slightly lower: 4.03 for the Cardiff Group and 3.82 for the South Africa and Yorkshire group. This implies that both groups felt that the show was more entertaining than it was educational.

If we look at the total numbers involved in the questionnaires and take an average of the percentage who selected a particular demonstration, then the top five cited as the 'most interesting' (along with the category of demo they represent) are shown in Table 4.

From this, we can see that looking at the most popular demonstrations *immediately* after the show, we have four incidences of a 'curious' category, three incidences of a 'human' category, two incidences of both the 'analogy' and 'mechanics' categories, and one incidence of 'phenomena'.

Data from the focus groups

As a follow-up to the questionnaires, two focus groups were held with the students who saw the show in December 2001 to try to establish what long-term memories remained. These took place in May 2004, and two groups of five students were invited

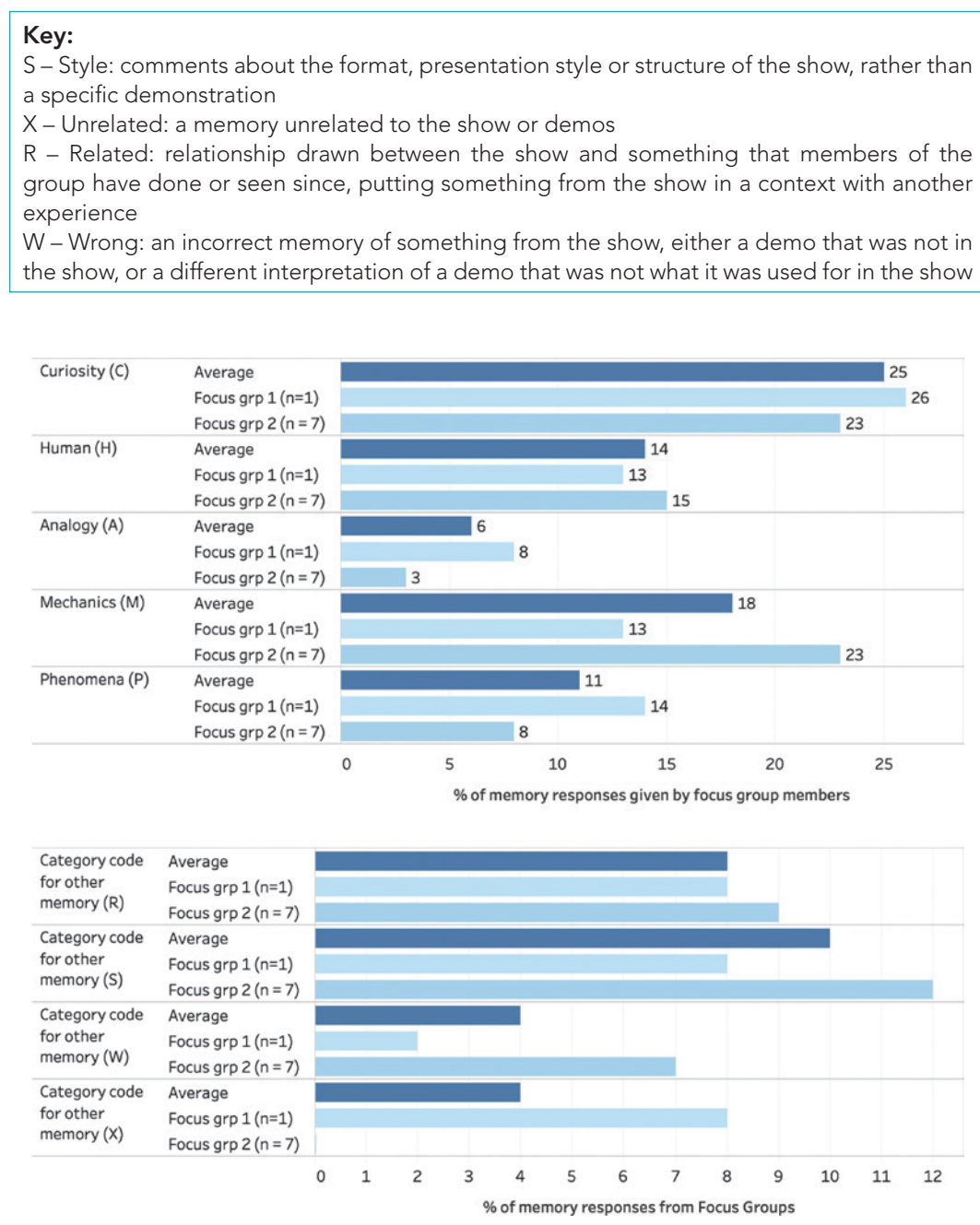
Table 4: 'Most interesting' demonstrations from short-term data

Demonstration	Categories used
1. Standing waves on a string (strobe light)	P, C, A
2. Theremin	C, M
3. Oscilloscope – seeing live sound waves on the screen	A, H
4. Voice synthesis (computer singing a Kylie Minogue song)	C, H
5. Changing a volunteer's voice	H, C, M

to attend on two separate days. Unfortunately, the days set by the school ended up being immediately after A-level examinations for some of the students. This means that the sample size of the focus groups is low, but it does provide some added depth to the other two data collection methods. From the findings of the questionnaires, a focus group script was written to guide the sessions. During the sessions, I used visual images and some props from the show to prompt memory recall. Figure 2 represents the data from the focus groups.

Where a memory occurred that related to a demo with more than one category, the initial coding of the memory was as *all* the categories associated with that demo. From then on, coding was assigned according to which aspect of the demo the student was talking about. When a memory was recalled that did not relate to the demos or

Figure 2: Data from focus groups



the presentation, it was coded with an 'X'. This was done so that the total percentage of memories recalled could be expressed in terms of specific demonstration memories and general other memories of the show.

The code 'R' was introduced to show that the person had made a connection between the demo and something else that was *related* but in another context. Statements that specifically recalled incorrect memories about the show were coded with a 'W' for wrong.

Focus Group 1 was the group consisting of just one student who had not studied A-level physics (the other four did not turn up). In this case, the 'curiosity' element of the demonstrations was the most frequently mentioned, at 26 per cent. 'Human', 'mechanics' and 'phenomena' were around the same level as each other, with 'analogy' being quite considerably lower. Another point to note is that there were six mentions of things that the student had related to other contexts since the show.

Focus Group 2 was the group of five A-level physics students and two teachers. In this case, there was an equal number of mentions (23 per cent) for 'curiosity' and 'mechanics' demonstrations, with the 'human' category in third place with 15 per cent. With this group, there were ten mentions of occasions where there was a relationship made with things since the show.

If we analyse the two focus groups together (giving us a total of eight people involved in this part of the research), then we see that 'curiosity' is the most commonly recalled category (25 per cent), followed by 'mechanics' (18 per cent), 'human' (14 per cent), 'phenomena' (11 per cent) and 'analogy' (6 per cent).

Finally, the analysis of how many different demonstrations were remembered in total throughout the focus groups is shown in Table 5. As there were 22 demonstrations in total in the show, it is interesting to note that by the end of the session, the students between them had remembered at least 50 per cent of these. Without any prompting, around 20 per cent of the demos were remembered.

Additional evidence from personal conversations and professional experience

An outline and summary of responses from my conversations with professional presenters was transcribed, and the same categories were used to code the quotations

Table 5: Total number (and names) of demos recalled throughout focus groups (out of 22 possible demos)

	With no prompts	Prompted by questionnaire	With pictures as prompts	With props as prompts	Total
Focus Group 1	5 Waves on a string, Computer singing, Hearing test, How sound travels, Oscilloscope	0	4 Shepard tones, Bucket and needle, Wine glass, Theremin	2 Slinky, Whirly tube	11
Focus Group 2	4 Whirly tube, Waves on a string, Computer singing, Resonance sticks	2 Digital and analogue, Theremin	6 How CDs work, Wine glass, Hearing test, Big ears, Shepard tones, Bucket and needle	2 Slinky, Whirly tube	14

to see which types of demonstration were mentioned by professional presenters, and how frequently.

From the qualitative data, we can see comments that reinforced the importance of the 'curiosity' factor. Presenters said:

... I always try and do something that will really surprise them ... (PP5)

Ones where the demos are counter-intuitive or different to what they expect are more memorable (PP2)

It is worth noting that with my categorization of demos, I included things that may not be thought of as demonstrations by other presenters. When asking the other presenters about demos, they were talking mainly about science demonstrations in their pure form, mainly showing an experiment or using scientific equipment. In this study, the aim was to encompass all practical ways of communicating a science message or phenomenon, which may include using volunteers and showing technology, but because the interviews should not be biased at all, this was not explained to them fully, and they were allowed to speak for themselves.

What the triangulation of research shows us

Despite some of the sample sizes being small, the triangulation of the three different types of data can give us some indicative results that may be applicable to further research. By comparing the results from the three different sets of data using the demo categories as a standard, we can investigate whether the initial impact, long-term memory and professional perspective about demo categories tie together (see Table 6).

Clearly, all the data agree that 'curiosity'-based demos have the most immediate impact, are most memorable in the long term, and are most important when professionals develop a show. In addition, the 'human' category comes in second on two occasions (and a close third on the other), suggesting that this is the next most important type of demo to both audiences and presenters. The only anomaly, where the presenters do not mention a category that is fairly popular both in the short and long term, is 'mechanics' (demos about 'how things work'). This was not mentioned at all in the presenter interviews, and yet it scores second and equal third most popular in the other data sets.

Table 6: Triangulation of data showing the prioritizing of demonstration categories from the different elements of the research

Questionnaires (initial impact) (%)	Focus groups (long-term memory) (%)	Professional presenters (presenter perspective) (%)
1. Curiosity (33)	1. Curiosity (25)	1. Curiosity (50)
2. Human (25)	2. Mechanics (18)	2. Human (25)
=3. Analogy (17)	3. Human (14)	=3. Analogy (12.5)
=3. Mechanics (17)	4. Phenomena (11)	=3. Phenomena (12.5)
5. Phenomena (8)	5. Analogy (6)	5. Mechanics (0)

Discussion

The short-term impact of a science show

Immediate reaction to the show

Many presenters will tell you of a real 'buzz' that is present at the end of a show. In my interviews with professional presenters, one said:

There is a huge difference when they enter the room and when they leave a room, and feedback I get from parents indicate that there is a great deal of activity that goes on afterwards. (PP6)

When you feel this reaction consistently with each presentation you give, it is perhaps not surprising that written feedback taken immediately after the event is overwhelmingly positive. The questionnaires from students who commented on the show in the short term showed that 100 per cent found that it was a positive experience. Nobody who was questioned rated the show as adequate or poor. The questionnaires were given to whole audiences and not just to selected members, so it is fair to assume that this is a representative result of the way most audiences feel after this kind of show. This alone could suggest that the science show is fulfilling its goal as a motivational experience. If 'education is about lighting fires, not filling bottles' (Knight, 2002: 217), then a well-presented science show can clearly be very successful at this – but does the effect last?

The results of this research showed a universally positive attitude immediately after the show.

Interest versus knowledge gained

The top five demonstrations that were said to be the most interesting were:

1. The strobe light (illuminating standing waves on a string)
2. The theremin (electronic instrument that you play by moving your hands in the air)
3. Oscilloscope (seeing live sound waves on the screen)
4. Voice synthesis (a computer voice speaking and singing)
5. The voice changer (using electronics to change the pitch of a volunteer's voice).

In contrast, when asked about one new thing that they had learned from the presentation, only the theremin was in the top five. The other things stated were more 'educational', and included general statements such as 'the relationship between physics and music'. This could mean that the most 'interesting' parts of the show are those that seem to be less 'educational', or, if we consider our demo categories, it ties into the fact that the novel and curious things are likely to be most memorable. By definition, these types of demonstration are probably not seen elsewhere, nor used in teaching in school, and because of this they may not be perceived to be educational. It is possible that the wording of these questions, both in the questionnaire and in the focus group, needs rethinking, as it introduces potential bias about what students think is meant by 'educational'. We have already changed these wordings on our other Science Made Simple questionnaires to ask about whether audiences 'learned anything new', as well as what they found most enjoyable or interesting. Further work could be done to explore whether the 'new things' are those that are most often remembered, to see if that links to the 'curiosity' factor of the demonstrations with the most impact.

In this case study, it was found that the content stated as being most interesting was not the same as that stated as being most educational.

The long-term impact of a science show

Which things were remembered after two and a half years?

The top five most 'interesting' things that were noted in the short term have a slight majority of 'curiosity' elements in them (33 per cent), and in the long term there is also a trend towards remembering the curiosity angle above all others. For example, the standing waves on a string demonstration fits within three category

types. Primarily, it illustrates a 'phenomenon' (standing waves), but in addition it is quite 'curious' (visually, it looks weird to see a string that appears to be frozen in space when viewed with a strobe light), and it is also an 'analogy' of what happens inside string and woodwind instruments when you play them. By careful analysis of the words used in the focus group, it becomes clear why the demonstration has stuck in the memory:

Because it was a *string*, and it was going like that, it was just cool!

... I didn't know you could do that with a string. (Focus Group 1)

On other occasions, there were similar statements made when asked about *why* certain demonstrations were memorable:

... yeah, that thing, it really stuck in my mind because I couldn't work out how you were doing it, because I play the piano and I remember I was mystified by that, but that was cool. (Focus Group 1)

I would have thought that the novelty of the theremin, like, made it far more interesting because people hadn't seen it before. (Focus Group 2)

... and how on earth could it work? (Focus Group 2)

Yeah, that sort of mysterious element. (Focus Group 2)

Looking at the average of both groups, the 'curiosity' type of demo had the biggest impact on the long-term memory.

Relating demonstrations to contexts beyond the show

The interesting thing about investigating long-term impact is that in the intervening time between event and evaluation, many occasions can arise that may cause someone to remember something that was said or done in the event. In addition to just recalling something that they have seen or heard, this means that they are adapting the knowledge to a different situation and, by doing so, it makes it more individual and real to them. In terms of the modes of memory from active and passive modes of memory listed above, we could assume that by making contexts for things they have seen, people are performing a mental manipulation with the information, and are therefore more likely to be able to recall it.

Although this was not the main focus of this study, there were a few incidences of that happening with the limited number of people used in the long-term study. In fact, around 9 per cent of all memories recalled involved relating something from the show to something that has happened in the time since the show. This may have been high because some of the students were studying A-level physics, and so were more likely to have come across similar phenomena in other contexts. However, a number of the 'related' memories were about situations beyond school. For example, the student who was not studying physics said:

Because I listened to my Dad playing a record just a couple of months ago, he doesn't play them often, but he has a lot of them, and he had forgotten to turn the speakers on, but if you put your ear right close to the needle you could hear it playing, and I was like, gosh, that is just the physical movement of a needle just pulling in a groove and it was like, wow ...!

Looking at the subset of memories that people related to other contexts, a range of demo categories are present. From the memories that people related to other things that had happened since the show (categorized as 'R'), six related to 'curiosity' or 'phenomena' type demos, five were about 'mechanics', three related to 'analogy' demos, and just one referred to a 'human' demo. It is perhaps not surprising that the 'analogy' and 'human' types are referred to less often in this context, as they are not things you are likely to come across in everyday life. The 'human' category requires an audience situation or volunteer to invoke the same reaction, and the 'analogy' is a tool for explaining that may not be seen again as being directly relevant.

Unlike in the other data, the 'mechanics' category scores highly in these 'related' memories. I assume that this is because these kinds of demo are about 'how things work', and therefore it is more likely that you will come across them again (or applications that are similar). Once again, the 'curiosity' factor scores highly. This is perhaps surprising, as a truly bizarre and 'curious' thing may not be something that is seen in day-to-day circumstances. However, perhaps the memory is so strong that people are more likely to look for other contexts for it, because it had a strong visual impact. 'Phenomena' demos also featured strongly in these 'related' memories and I suspect this is because the physics phenomena shown are very likely to be shown again in different ways within the classroom teaching of physics. As one focus group had studied A-level physics it is more likely they would have seen related phenomena through their school learning.

These results suggest that 'curiosity', 'phenomena' and 'mechanics' demos are most likely to be the ones remembered and applied in related memories over the longer term.

Differences between the physics students and the non-physics student

Despite there only being one student in the non-physicist group, she remembered more material unprompted than the five students and two physics teachers together. This surprised me, but in fact it is probably to be expected. If (as suggested in this research) we are proposing that novel activities are more memorable, then it is likely that all the physics demos were novel to her, and perhaps have not been repeated anywhere since. However, those who studied physics are likely to have come across versions of the demonstrations through their studies, and perhaps this means that they are less 'novel' or 'curious', and therefore less memorable. However, the very low numbers mean that it is hard to draw any significant conclusions from these comparisons.

Are the short-term and long-term impact related?

Of the top five demonstrations stated as being the 'most interesting' in the short term, three of them also came up as unprompted memories in the focus groups two and a half years later. This suggests that the favourite demo in a show is more likely to be remembered in the long term than something which does not appeal immediately after the show. However, the theremin (which was the second most popular demo from the questionnaire data) did not show up at all as an unprompted memory. Instead, the resonance sticks and the test of the hearing of the audience were mentioned as unprompted memories, and these were not mentioned as some of the most interesting sections of the show. As the theremin is such a bizarre and abstract instrument, it is likely that this may not have been seen in any context since the show, and so was easily forgotten. However, on seeing a picture of the instrument, most of the focus group students did recognize it and could tell me something about it. In a small-scale sample where a total of seven demos are mentioned unprompted in the long term, it is hard to judge whether three of them being the same as the short-term data is significant.

Which types of demonstration have the most impact?

Looking at all the data available, it seems that by some considerable majority, demonstrations that are curious, novel, counter-intuitive or involve a challenge about the outcome are most commonly recalled. Without using any of the category descriptors, there was a remarkably similar phrase used by the student in the first focus group and one of the professional presenters when talking about a very popular demo. When asked why they liked this particular demo, they said:

... because it was a *string*, and it was going like that..., it was just cool
(Focus Group 1, ref 11)

Regarding a different demo where a gherkin glows orange like a sodium lamp:

... because it was a *gherkin* with mains electricity going through it! (PP4)

Both were said with a tone of voice that suggested 'well, it's obvious it was great because it was just so bizarre!'

The results also show us that the 'human' angle (using a volunteer, finding out personal information about yourself, or something funny with a member of the audience) is highly rated as well. The human angle is of course closely related to the fact that this is a live theatrical performance that uses people as volunteers. Sometimes the only way to involve the whole audience directly is to get all of them doing an experiment together, and this is a widely used technique. In addition, humans are essentially egotistical animals in that they like to learn something new about themselves. In this case, the test of your hearing or the fact that your ears can be fooled tells you something personal to you, which has more impact than learning about an object.

The 'mechanics' category was the only area where the opinions of the professional presenters differed greatly from what is suggested by the focus group and questionnaire data. The focus group data suggest that these demonstrations are very memorable, although this may be skewed by the proportion of A-level physics students who took part. However, as the focus group data show, these memories were particularly with reference to things that they had related to since the show. This could be a recommendation for good practice, as making links with other things is a good sign that there has been some positive and long-term impact.

It is perhaps surprising that the pure 'phenomena' demonstrations (showing a science phenomenon as it happens) are less popular, as these have usually been thought of as the 'bread and butter' of these shows. This could be because many show presenters are now using a 'phenomena'-based experiment, but interpreting it in such a way as to make it 'curious' or surprising. For example, the demo with the gherkin shows how a sodium solution can be made to glow orange when energized. In the past, this would have been demonstrated using a sodium tube plugged into a light socket, or perhaps using a picture of old-fashioned streetlights. The gherkin demonstration shows the same science, but in a bizarre way. The gherkin is hooked up to mains electricity (don't try this at home!), and because of the high salt content of the pickle, it glows orange. This is probably a sign of the times, as we search for ways to engage ever more sophisticated audiences who may have seen the basic phenomenon presented many times before online or on television. The 'curious' demo often retains elements of a 'phenomenon', of course, but the overwhelming aim of it is to make people surprised or amused, rather than just to witness the phenomenon in its own right.

The 'analogy' type of demonstration were also not high on the priority list. Presenters are very aware of the use they have for explaining a concept, but they are

primarily an educational tool, rather than a motivational one. It is not surprising that demos used as analogies are not particularly memorable on their own.

This study provides some evidence that the 'curiosity' and 'human' types of demo can have the most significant long-term impact.

Why does the 'curiosity' type of demo have the most impact?

Something that is unusual and unexpected raises your awareness the instant it happens. In psychological terms, this is referred to as cognitive dissonance, and research has shown that this can lead to a 'drive' to resolve that difference (Beswick, 2017). This could explain the memorability of this 'curiosity' type of demonstration. There is a need to resolve what appeared counter-intuitive, which makes the demonstration more memorable and more applicable as the viewer tries to make sense of what they saw.

In addition, if you are trying to recall something from a long time ago that uses everyday equipment in an ordinary way, then your memory may be blurred with many other memories of similar uses of that equipment. However, if an everyday item is used in a bizarre way (for example, making a musical instrument from a drinking straw or making a gherkin light up) then you will not have too much trouble distinguishing that from other uses of that item.

Science centre exhibit research has also provided some evidence that the 'novelty' (or 'curiosity') level of an exhibit has an effect on its memorability, and possibly even the ability to encourage cognitive processing (De Witt and Osborne, 2010). In addition, functional magnetic resonance imaging (fMRI) has been used in recent years to show how curiosity can increase the ability to remember items (Gruber et al., 2014).

Limitations of the research

It is acknowledged that the research has a number of limitations. Due to the fact that the research concerns a show that was developed and presented by the author, it would be naive to suggest that the analysis of the results could be completely objective. In addition, it is likely that the author brings to the work opinions about shows beyond the one being analysed.

It should also be noted that because the school used for the focus group was a single-sex school, the focus group data are all from female students. Some research suggests that male and female secondary school students have different reasons for finding physics interesting. Boys are engaged by the practical work, whereas girls report more interest in the way physics relates to everyday things (Williams et al., 2003). The questionnaires (which shaped the focus group questions and the choice of demo categories) did come from mixed-sex audiences.

Conclusions and recommendations

Summary of conclusions

This research has provided a rare opportunity to compare the short-term and long-term impact on audience members. Short-term quantitative and qualitative data were collected, and two focus groups were held to establish long-term impact and devise a set of demo categories to help analyse the data. Finally, these results were enriched by triangulating the data by interviewing professional presenters.

In summary the research found the following:

- It is useful to develop a framework of categories as an analysis procedure for the effect of science demonstration shows. In this study a framework called CHAMP

was introduced made up of five demo categories: Curiosity, Human, Analogy, Mechanics and Phenomena. Some demos have more than one category element, but all will fit somewhere within these definitions, and most have a clear primary category which can be defined as the reason the demo has been selected at that point in the show.

- Without any verbal or visual prompts, members of the focus group managed to recall around 25 per cent of the demonstrations used in the show after a period of two and a half years had elapsed. With some visual prompts, the groups managed to recall over 50 per cent of the demos used.
- There is evidence that members of the audience have made related links to things they saw in the show in other contexts. Around 9 per cent of the memories from the show in the long term related to this kind of recall.
- The 'curiosity' type of demo (bizarre, novel, unexpected, counter-intuitive, challenging) consistently comes out as the demo type that has the biggest impact – in that it is most memorable, most commonly referred to in conversation and most frequently cited by professional presenters when talking about the types of demo they like to use.

This study has been incredibly useful for professional development as an informal science learning practitioner. After more than twenty years of actively writing and presenting shows, there has not previously been the time or opportunity to review what is being done beyond a 'snapshot' of evaluation immediately after the show. I have been surprised by some of the results (that 'phenomena' demos have fairly low impact), while other data have reinforced my instincts about what works particularly well.

There was a higher than expected amount of recall from the focus groups two and a half years after the show, which is quite inspirational. This backs up the suggestion that people do recall demos some time after a show is over (Burns, 2003). One of the most exciting things for me was to hear that the students had made links between things they saw in the show and science in other contexts. There is sometimes criticism that events such as science shows and festivals have a short-hit lifetime that is quickly forgotten. If this research suggests that even one or two things remain with someone long enough for them to process the memory and use it in an applied situation, then this is a real achievement for informal science learning.

Guidelines for best practice

Based on this research, it is recommended that science show professionals ensure that there is a mixture of the CHAMP demo categories in their presentations, as there is evidence to suggest different types of audience respond to different categories of demonstration. However, this research suggests that some generalizations can also be made:

- 'Curiosity'-type demos seem to be universally popular regardless of the audience, and they have a high impact rate for short-term and long-term recall.
- 'Human' demos are also highly memorable, so they also have a fairly high impact.
- 'Mechanics'-type demos are more popular with audiences that are already interested in science. 'Mechanics' demos are also the type that are most likely to help people relate the show to other contexts.
- 'Analogy' and 'phenomena' demos are useful educational tools, but they tend to have less short-term and long-term impact.

In addition, the data suggest that short-term impact is likely to be similar to that which is remembered in the long term. This potentially means that when there is a lack of resources for longitudinal studies, it may be possible to extrapolate from the short-term impact to make a hypothesis about the kind of things that are likely to be remembered over a longer period of time.

Opportunities for further research

The scale of this project has not allowed full exploration of all the elements of the research data. The author would encourage others to use this model as a basis for further exploration, and welcomes ideas for future collaborations to advance the field. Possible projects that would help build on this knowledge include:

- applying these demo categories to a variety of shows and a variety of different aged audiences to see if different category types really do tend to appeal to different ages
- trying short-term, medium-term and long-term research to see how the memories change over time
- conducting a similar longitudinal study, but trying to include some data about attitudes rather than just the memories; if the affective domain is where the science show aims to make a difference, then this needs to be measured in some way.

Acknowledgements

This research was conducted as part of an MSc in science communication from the Open University, UK – without whom I would not have been able to study further while keeping the job I loved as a practitioner. I would like to thank the Open University for this opportunity to support practitioners in gaining research skills alongside their work.

Notes on the contributor

Wendy Sadler is a senior lecturer in science communication at the School of Physics and Astronomy, Cardiff University, UK. She is also the founding Director of *Science Made Simple* – a social enterprise and Cardiff University spin-out (established in 2002) with a mission to inspire the next generation of scientists and engineers.

References

- Beswick, D. (2017) *Cognitive Motivation: From curiosity to identity, purpose and meaning*. Cambridge: Cambridge University Press.
- Bultitude, K. and Eigenbrot, I. (2004) *Lasers: Light up your life. Evaluation report*. Unpublished report.
- Burns, T. (2003) *Science Shows: Evaluating and maximising their effectiveness for science communication*. Newcastle, NSW: University of Newcastle.
- DeWitt, J. and Osborne, J. (2010) 'Recollections of exhibits: Stimulated-recall interviews with primary school children about science centre visits'. *International Journal of Science Education*, 32 (10), 1365–88. <https://doi.org/10.1080/09500690903085664>.
- Di Stefano, R. (1996) 'Preliminary IUPP results: Student reactions to in-class demonstrations and to the presentation of coherent themes'. *American Journal of Physics*, 64 (1), 58–68. <https://doi.org/10.1119/1.18293>.
- Gruber, M.J., Gelman, B.D. and Ranganath, C. (2014) 'States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit'. *Neuron*, 84 (2), 486–96. <https://doi.org/10.1016/j.neuron.2014.08.060>.

- Herrmann, D. and Plude, D. (1994) 'Museum memory'. In H. Falk and L. Dierking (eds), *Public Institutions for Personal Learning: Establishing a research agenda*. Washington, DC: American Association of Museums, 53–65.
- Hussey, J. and Hussey, R. (1997) *Business Research*. London: Macmillan.
- James, F.A.J.L. (2002) "'Never talk about science, show it to them": The lecture theatre of the Royal Institution'. *Interdisciplinary Science Reviews*, 27 (3), 225–9. <https://doi.org/10.1179/030801802225003178>.
- Knight, D. (2002) 'Scientific lectures: A history of performance'. *Interdisciplinary Science Reviews*, 27 (3), 217–24. <https://doi.org/10.1179/030801802225003277>.
- McManus, P. (1993) 'Memories as indicators of the impact of museum visits'. *Museum Management and Curatorship*, 12 (4), 367–80. [https://doi.org/10.1016/0964-7775\(93\)90034-G](https://doi.org/10.1016/0964-7775(93)90034-G).
- O'Brien, T. (1991) 'The science and art of science demonstrations'. *Journal of Chemical Education*, 68 (11), 933. <https://doi.org/10.1021/ed068p933>.
- Stevenson, J. (1991) 'The long-term impact of interactive exhibits'. *International Journal of Science Education*, 13 (5), 521–31. <https://doi.org/10.1080/0950069910130503>.
- Stohr-Hunt, P.M. (1996) 'An analysis of frequency of hands-on experience and science achievement'. *Journal of Research in Science Teaching*, 33 (1), 569–600. [https://doi.org/10.1002/\(SICI\)1098-2736\(199601\)33:1<101::AID-TEA6>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1098-2736(199601)33:1<101::AID-TEA6>3.0.CO;2-Z).
- Williams, C., Stanisstreet, M., Spall, K., Boyes, E. and Dickson, D. (2003) 'Why aren't secondary students interested in physics?'. *Physics Education*, 38 (4), 324. <https://doi.org/10.1088/0031-9120/38/4/306>.
- Wu, H., Krajcik, J.S. and Soloway, E. (2001) 'Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom'. *Journal of Research in Science Teaching*, 38 (7), 821–42. <https://doi.org/10.1002/tea.1033>.