Duca E. et al. (2022). *Xjenza Online*, 10(2):86–102.

Xjenza Online: Science Journal of the Malta Chamber of Scientists
www.xjenza.org
DOI: 10.7423/XJENZA.2022.2.03

Review Article

Designing an evaluation strategy for a large-scale science and arts festival using Science in the City, Malta as a case study

S. Cutajar[1], **S. Seligova**[2], **E. Duca***[3]
[1] *Institute of Earth Systems, University of Malta, Msida, Malta*
[2] *DLHA38, Stupava, Slovakia*
[3] *Department of Mathematics and Science Education, Faculty of Education, University of Malta, Msida, Malta*

Abstract. In this work, we analyse and present a step-by-step guide on how to set up a valuable and informative mixed method evaluation strategy of large-scale science festivals and events. A literature analysis helped identify the best technique to set up a multi-approach methodology (multiple-choice questionnaire and silent observers). Questionnaire data was to be collected using systematic sampling. The approach was applied to a local case study to develop best practice. Its implementation was analysed and assessed to provide festival organisers with useful recommendations to enhance the evaluation strategy, and improve festival quality and researcher engagement in subsequent editions. Combining a mixed-method approach to collect both qualitative and quantitative data helped gather a good and comprehensive overview of the festival. It set a baseline for future editions to improve upon. All the evaluation efforts carried out in this work were very dependent on volunteers, therefore an effective and appropriate volunteer recruitment, training and retainment strategy was essential. This work has developed a baseline assessment of the festival establishing a professional evaluation strategy with limited funds and experience. It is a step-by-step guide for large science festival organisers who want to set up effective evaluation of their efforts.

Keywords: Public engagement with science and technology, Informal learning, Science festival, Evaluation strategy, Qualitative data, Quantitative data

1 Introduction

The phenomenon of large public science communication events such as science festivals has spread throughout Europe and the world (Bultitude et al., 2011) with a dramatic increase in the last few decades (Cassidy, 2006). Science festivals are time-limited and recurring celebrations of science that engage non-specialists with scientific content using activities with common themes and branding (Bultitude et al., 2011). They include a large variety of events showcasing STEM subjects through lively and entertaining events including hands-on activities, live experiment demonstrations, exhibitions, workshops, and lectures. European Researchers' Nights, supported by the European Commission's Research and Innovation Framework Programme Horizon 2020 (H2020, 2014–2020) by the Marie Skłodowska-Curie actions, are public events dedicated to bringing researchers closer to various publics. Under the criteria set by Bultitude et al. (2011), European Researchers' Nights are considered science festivals and will be treated as such in this work. Science communication researchers argue that science festivals are ways to communicate science between scientists and different publics (Burns et al., 2003; Jensen, 2014; Jensen & Buckley, 2011, 2014). Only a few, however, have systematically and consistently examined the effect of science festivals on learning, practicing and inspiring science, improving attitudes towards science or enhancing scientific literacy (Burns et al., 2003; Jensen & Buckley, 2011). Similarly, there is little published research available on the evaluation of European Researchers' Night, with Dimitrova (2010), Roche et al. (2017) and Sardo (2016) being rare examples. In fact, science festivals have long been criticised for their lack of rigour when evaluating (Bultitude et al., 2011) and consideration of long-term research towards their impact (Jensen & Buckley, 2014), both of which require methodological expertise and resources by the event organisers that are not necessarily available (Jensen, 2015; King et al., 2015). Jensen (2014) and Jensen and Buckley (2014) pointed out that the methodologies of some studies related to science festivals are prob-

Correspondence to: E. Duca (edward.duca@um.edu.mt)

lematic. They are heavily dependent on close-ended and simple survey data, so that complex and insightful feedback is often lacking. Consequently, analysing and evaluating the multitude of dynamic events happening during a science festival is often seen by organisers as a burden and a box-ticking exercise undertaken for funders. Effective science communication can empower research and innovation systems to address global challenges and put public interests at the heart of how knowledge is produced, shared and applied. However, for science communication to play this mediating role effectively, there needs to be a more integrated and 'evidence-based' approach (Jensen & Gerber, 2020). As such, practical guides on evaluation strategies of science festivals are useful tools for organisers to start evaluating their impact in an appropriate way. This paper is a step towards evidence-based science communication practice. It offers a step-by-step guide to support science communicators and organisers with little or no experience in evaluation. It outlines the process of evaluating large festivals that involve a variety of activities by using a mix of qualitative and quantitative data.

2 Context

2.1 Case Study: Science in the City, Malta

This paper highlights a science and arts festival on the Maltese Islands, Science in the City (SitC), as a case study for designing and conducting an evaluation strategy of a large-scale science festival. Science in the City, set up in 2012 partakes in the EU-wide European Researchers' Night, and is the Maltese Islands' most high profile annual science activity. The festival is supported by the European Commission's Research and Innovation Framework Programme Horizon 2020 (H2020, 2014–2020) by the Marie Skłodowska-Curie actions. At its sixth edition, the festival was estimated to have attracted more than 20,000 people to Valletta, the Maltese Island's capital city (over 5% of the Islands' population). Led by the University of Malta, the primary aim of the festival is to engage with and involve local people within a variety of science themed activities. The festival marries research, mostly undertaken in Malta, with theatre and dance performances, stand up comedy, art exhibitions, art installations, hands-on activities, music and science experiments. In its sixth edition, the festival showcased six main science and arts-based activities and opportunities, varying from jazz music, interactive artistic installations, music, theatre and dance. Science in the City (SitC) has established itself as the leading exemplar of best-practice in arts-science engagement in Malta. Utilising SitC, a best practice strategy to evaluate a science festival was tested and documented. Current literature on science festival evaluation and large touristic events was reviewed and used to develop a comprehensive strategy. The multi-method evaluation strategy developed and tested in this research was intended to serve as a basis for adaptation by other event evaluators. An implementation plan is outlined, including step-by-step methodology and management of human resources.

2.2 Evaluation Strategy

An evaluation plan to explore the strengths and weaknesses of SitC as a public engagement event was designed. A number of research methods were considered for the evaluation, such as participant interviews, visitor tests, archival document review and case studies (Shaw et al., 2006). These methods were considered unsuitable or impractical for a public engagement event on the scale of SitC. For example, participant interviews during the event would distract from the fun, entertainment value of visitors, while archival document reviews were impossible as previous evaluation data was not of sufficient quality. The goal was to capture as much information as possible across the breadth of events within a tight time-frame. A mixed-method evaluation approach consisting of surveys collected using systematic sampling (every fifth person observed by the trained surveyor in specific age groups) and independent observations by silent observers was chosen.

3 Evaluation objectives

The specific objectives of the festival's evaluation were identified following the 'SMART' principle, i.e. specific to the event, measurable, achievable (or attainable), realistic (or result-oriented), and time bound around the event schedule. The objectives were targeted for this case study of an annual large science and arts festival that is attended by tens of thousands of people. As this was the first attempted systematic evaluation exercise of this festival, efforts were limited to the following objectives which were identified as baseline information that an organiser of a science festival would want to understand as an initial exercise:

1. Demographics: Audience demographics profile, including gender, ethnicity, income, country/town of origin and level of education (Questions 1–5, Appendix A).
2. Profile of visiting group: Visiting group characteristics, including group size and makeup (eg. are people visiting with other adults or children? Questions 6–9, Appendix A).
3. Motivation for attending: Understanding the attendee's reasons in visiting a science and arts festival (Question 10, Appendix A).
4. Quality of experience: Several questions were asked about the nature and quality of experiences at the

festival, including whether they enjoyed the festival or not (Question 11, Appendix A). Further questions assessed how comfortable attendees were with the interactions with researchers and their level of participation, to understand if researchers were engaging with visitors in an effective way (Questions 12–16, Appendix A).

4 Methods

4.1 Mixed-method evaluation

A mixed-method approach (Seaton, 1997) was chosen to be the most suitable for a large festival with activities ranging from artistic expressions to science experiments, outdoor to indoor locations, and professional performers to university student volunteers. The mixed-method approach adopted consisted of:

1. A 16-item multiple-choice questionnaire (see Appendix A). Questions were adapted from the literature to ensure validity and reliability. It included questions on visitor profiles (demographics, accompanying group size and composition, education level, and place of residency), motivations for participation, reception of the festival, and interactions with researchers/volunteers. The questionnaire was administered through personal interviews by trained volunteer interviewers who sampled the population in a systematic manner. Both online and paper versions of the questionnaire were provided to interviewees according to their preference.
2. Silent Observers (see Appendix B): three trained volunteers were instructed to experience the festival from a visitor's perspective. These observations were noted in a structured manner. Each silent observer was given a list of activities and events to attend and different starting points and places to be throughout the night.

4.2 Respondents of the questionnaire

The quantitative evaluation was conducted on the night of SitC on 30th September 2016. A total of 392 entries were recorded over the course of five hours. This number was limited to audience members. The questionnaire was printed on A4 paper in English or on an online platform.

4.3 Volunteers recruitment, training, and management

Bussell and Forbes (2002) highlighted the importance of understanding the target audience in order to successfully recruit and retain volunteers. The evaluation volunteers were undergraduate and postgraduate students from the University of Malta and employees of a pharmaceutical company (as part of their Corporate Social Responsibility) that had shown interest to help out with the festival. In this way, it was ensured that there was a diverse sample of students and professionals participating in data collection. When recruiting students for the evaluation strategy, preference was given to those who were unaffiliated with an organisation since it was expected that students that form part of student organisations would prefer volunteering with their peers. Students were specifically sought out in a dedicated call for volunteers that was advertised through the main channels used by students (such as Facebook). Finally, for the silent observers, volunteers who the organising team personally knew were relied upon due to the level of trust required for the task. The importance of understanding what motivates people to volunteer for the festival's evaluation was acknowledged in order to offer better support and guidance (Hyejin & Ross, 2009). This was achieved by meeting every volunteer a few weeks before the festival. A two-hour workshop introduced volunteers to the festival's evaluation practice, the importance of good data collection and unbiased sampling, and the event details and logistics. Volunteers had the freedom to select which age category they would survey during the festival, ensuring they would feel most comfortable when talking to interviewees. Management ensured that an equal number of volunteers covered each age category. Suggestions on how to select respondents were illustrated and practiced: volunteers were instructed to first look down at their feet, look up for people in their selected age category, count five such people, interview the fifth person, and, if the selected person would not agree to take the survey, start counting again—a systematic sampling approach. During the workshop the online survey was installed and tested on their smartphones (see Logistics). Interviewers received a free festival t-shirt and an 'Evaluation staff' name tag to associate them with the festival to aid approaching festival attendees. Free water and snacks were provided. Additionally a symbolic reward (€10 vouchers from a local book shop) was given to the participants at the end of the night's work. The reward was only mentioned after they had undertaken the necessary training and showed up on the day. This was done to incentivise future participation and to thank them for their work during the night. The volunteers were all given basic training on approaching citizens. Sample scripts were provided to help them introduce themselves as evaluation volunteers with SitC, give details of the research project and invite festival attendees to take part in the survey. Science in the City lasted for approximately six hours and the volunteers worked individually for nearly five hours of the event (one hour after the event started till half an hour before its end). They were also advised on how to choose

participants randomly to reduce sampling bias based on an individual's gender, age or ethnicity. The volunteers used paper-based versions of the survey or digital version on tablets or other smart devices, depending on their and the interviewee's preferences. The volunteers were also located in strategic locations in the festival such as 'rest areas', 'meeting points', 'entrances' and 'exits' in order to observe public reaction and participation in the events.

4.4 Quantitative data collection

4.4.1 Questionnaire design

The questionnaire was designed in English with the intention of capturing a diverse range of responses in 10 minutes or less per survey participant. Following published guidelines on designing questionnaires for evaluations (Boynton & Greenhalgh, 2004; Spicer, 2012) the survey relied on a mixture of several closed-end multiple choice questions and limited open-end questions (Reja et al., 2003). The survey questions were adopted from reports and guidance on evaluation of public engagement activities that investigated similar ideas at different public engagement events. Apart from demographic and general questions (Questions 1–5), the questionnaire consisted of questions related to motivation of attending SitC (Question 10), level of participation (Questions 12, 16), knowledge and interest in science (Question 9) and impression of the festival (Question 11) . The draft questionnaire underwent pilot testing amongst the SitC committee and volunteers to resolve the issues of ambiguity and incomprehensibility until it achieved a form that gathered responses on the most pertinent points of the evaluation.

4.4.2 Questionnaire platform

Although volunteers were recruited to interview people and gather data in person, a system to collect entries was still necessary. An electronic data collection methodology was opted by the research team for three main reasons. First, the technology provided a reusable system for future festival editions allowing for on-going evaluation. Second, the collected data did not need to be converted into an electronic format for analysis avoiding extra labour costs and reducing risks of introducing additional errors. Third, the ease of data entry and logistics: volunteers could access the survey at any moment from their smartphones without any limits on the number of interviewees. The software package ZohoForms was identified as the best online survey tool after a gap analysis. The analysis looked for affordable and robust software with both online and offline data collection, and with the option to be installed on multiple smartphones and operating systems. Only a few volunteers expressed a strong preference towards collecting data in a paper format, which was accommodated.

4.4.3 Data management and analysis

All collected data were collected in a digital or paper format. Data was entered into a Microsoft Excel database, formatted and homogenised through the use of the Comma Separated Values (CVS) format and analysed using R programming. The descriptive analyses included descriptive statistics such as frequency, mean and percentage. Inferential analyses were made to calculate the total median score on their overall ratings using a Likert scale on the following variables (Questions 9, 11, 12): 1. How do you rate your general level of interest in science **outside** of the Science in the City festival? (score range 0–4); 2. What is the overall impression of the 2016 Science in the City? (0–4); 3a. To what extent do you agree or disagree with the following statements? (0–4) 3b. I am likely to further investigate some of the scientific topics I encountered at the SitC festival (0–4). The calculated overall score derived from the Likert scale questions was treated as a dependent variable. This was termed as interviewee score. Median values of interviewee score were compared between various categories of participants, treated as independent variables, for example age groups, gender, and 'yes' or 'no' responses to questions. Mann Whitney U and Kurskal Wallis tests were used to test the differences in median scores. The difference in mean and median scores were statistically considered significant if p-value was less than 0.05.

4.5 Qualitative data collection

4.5.1 Silent observers

Following Seaton (1997), an unobtrusive observation method was adopted to record and study the events. The three silent observers were instructed to dress normally and to show minimal signs of their role at the festival in order to reduce possible positive or negative biases. Each volunteer was given a notepad, a pen, a map of the festival and the full programme. To ensure that the organisers could have an overview of the entire festival, a suggested schedule of events to attend and places to visit was provided. However, silent observers were instructed to move freely, observe and experience the festival. They were specifically asked to keep notes on: logistics, audience, and the behaviour of other volunteers and researchers (see Appendix B).

4.5.2 Crowd estimates

The most challenging aspect of the study was estimating the number of people attending the festival. Management discussed and researched multiple methods of crowd counting but did not find an optimal solution as it was

either too human resources dependent, using clickers at entry points; or too expensive/technical, using algorithms and images of the crowd (Cariveau, 2014; Streich et al., 2003). Management decided to use crowd estimates given by the local council and police. However, due to the high mobility of people and gatherings in front of shows, exhibitions and activities happening throughout the night it was very difficult to estimate the density of people along the streets and in indoor spaces.

4.5.3 Minimum number of volunteers

From previous editions of the festival, a minimum of 20,000 people were roughly estimated to attend the event throughout the night. A statistically significant sample at 95% confidence level (with 5% confidence interval) requires about 377 responses from the questionnaire. We needed a minimum of 8 volunteers collecting approximately 10 surveys an hour for the five hour long surveying period (or 50 surveys per volunteer over the night). In order to limit biases and to make sure that all age categories were well represented, the population of festival goers was divided into four age categories: 0–17, 18–29, 30–49, and above 50 years old. Each volunteer was instructed to gather data of people in one age category only. An equal number of volunteers was assigned to each age group.

4.5.4 Logistics

On the night of the festival all volunteers met with the evaluation manager half an hour before the start of the festival. At this point all materials were distributed and last minute directions were given. Data was gathered from about one hour after the event started till half an hour before its end. Half an hour before the closing of the festival a debriefing meeting took place: all interviewers and silent observers gathered and discussed with the evaluation manager their impression of the festival, the data collection methodology and other feedback.

5 Results

5.1 Results from the case study

Nineteen volunteers collected 392 survey responses during the festival. The sample was smaller than the study's initial objective. The reason is due to an overestimation of the working capacity of volunteers.

5.2 Demographics

Despite the sample's small size, the data collected is well-sampled: results show an almost equal number of males and females were surveyed (48.7% female; 51.3% male); and respondents are also equally split across age ranges as shown in figure 1 (blue represents direct sampling, 392 responses). figure 1 also illustrates how the group demographics data gives a qualitative idea of the real population present at the festival (orange represents group sampling, 1,237 indirect responses). From such data emerges that the bulk of participants are below 15 years old, and between 16–25 years old.

Figure 2 shows two important survey results: (1) mostly families (more than 50% of surveyed people) visit the festival, and (2) there is a considerable portion of young adults (16–25 years old) attending the festival with friends. Independently, silent observers reported a lack of activities and experiments targets towards young adults, but plenty of experiments for children. Due to the large attendance from young teenagers, future editions of the festival should try to specifically target such an audience.

5.3 Reception of the festival

Forty-two percent of the people surveyed were going to the festival for the first time. 83% of respondents liked the festival. Two of the silent observers had previously experienced the festival. Their comparison with past editions could identify practices that improved (or not) and gave feedback on possible improvements. The point of view of the other silent observer that never participated in the event was also relevant to point out what a newcomer notices, appreciates, and dislikes.

5.4 Reason for attending

Of the sampled population 76.5% have an interest in science and 59% of them would investigate science further after the festival. Reasons for attending are mainly to have fun (39%), to spend time with the family (26%), and to find out more about University of Malta research (21%).

5.5 Interactions

While 88% of survey respondents felt that researchers or volunteers at stands were enthusiastic when communicating to attendees, only 54.6% approached them to engage in conversation with them, and 14.6% felt that they were using difficult jargon. As the night progressed volunteers engaged less with visitors.

5.6 Internal feedback

The three silent observers collected qualitative data on the festival with special attention to one-off events, performances, and indoor activities. From their notes, it emerged that the composition of festival attendees changed throughout the night. Their feedback turned out to be an important internal feedback mechanism that gave insight on logistics, hard places to find, scheduled events not run on time, low attendance in certain events, and practical information. Table 1 summarises the multi-

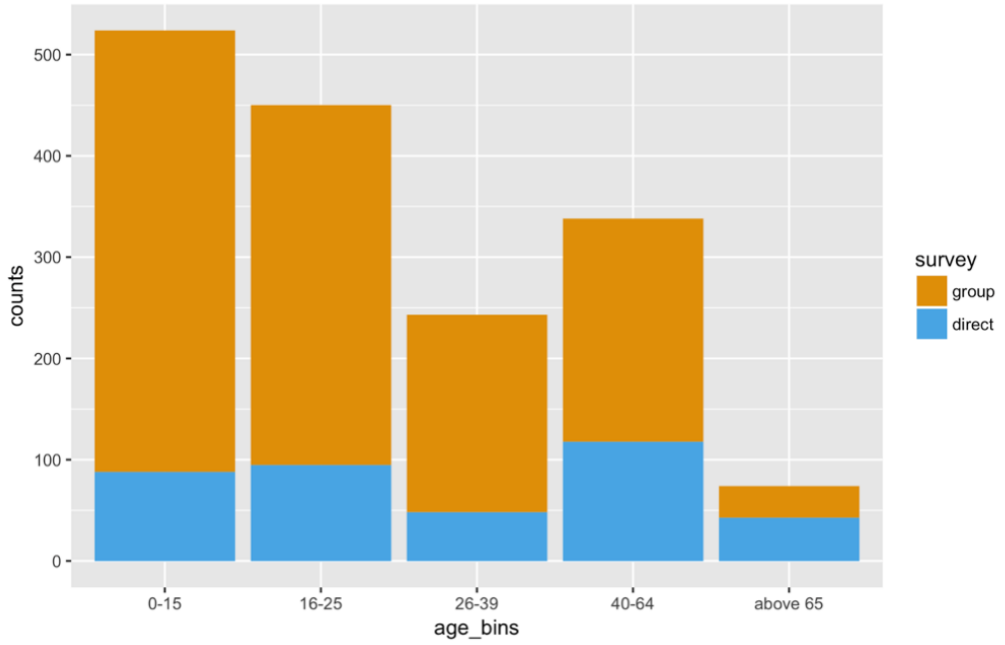

Figure 1: Age spread of all surveyed people (1629 in total), both directly (total of 392 people in blue, bottom) and indirectly through group demographic questions (total of 1237 in orange, top).

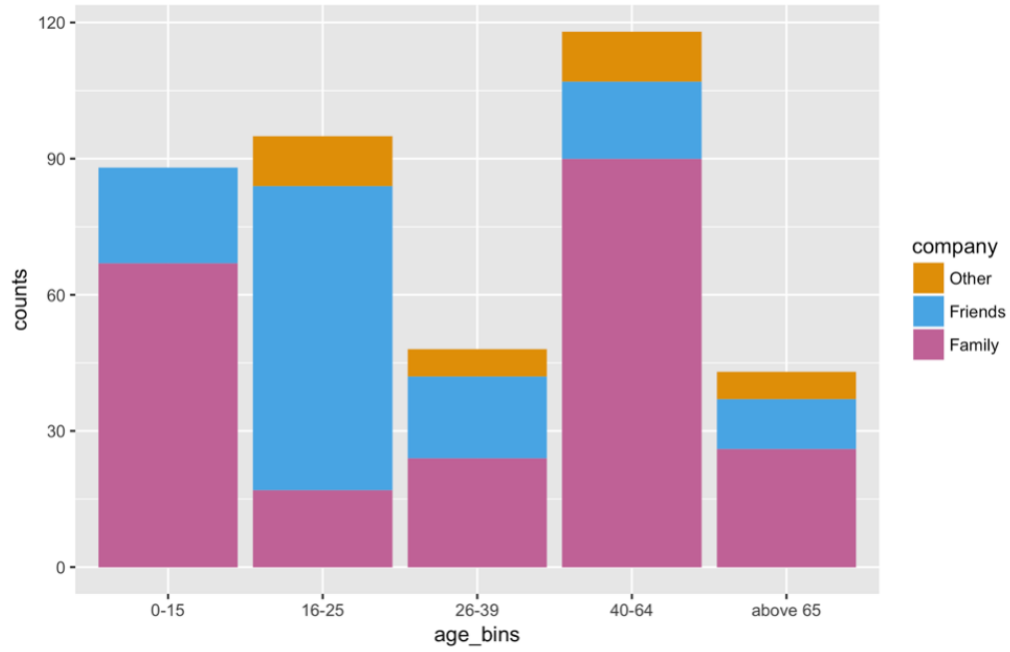

Figure 2: Age spread of directly surveyed people (392 in total) with indication of their company (Family, Friend or Other).

strategy evaluation and its utility according to the dimension that was researched and analysed.

6 Discussion

Overall, setting up a mixed method evaluation strategy was not as challenging as anticipated. Good results were achieved in both methodology and information gathered. However, a few issues need further investigation. In terms of methodology, more attention must be paid on volunteer recruitment and management in order to optimise efforts from the organisers.

6.1 Setting up the evaluation

The time needed to create the evaluation strategy was considerably less than expected due to two main factors:

1. extensive research has already been carried out on large touristic events (Langen & Garcia, 2009; Wood, 2005) which provided adaptable guidelines for science communication practice (del Carmen Sánchez-Mora, 2016; Koolstra, 2008);
2. case studies and examples of survey questionnaires from similar events were publicly available and easily adaptable. Instead the time of the research team was mostly invested in recruiting, training, and managing volunteers.

6.2 Volunteer strategy

An appropriate and effective volunteer recruitment and management strategy is essential for the success of a large-scale evaluation strategy. The volunteer recruitment process had its challenges. Although the organising team started contacting organisations two months in advance, potential volunteers only started responding three weeks before the festival. The main contribution of volunteers came from a medical research company that had an interest in helping out with the festival. Only when the organising team successfully identified a key contact person did all the communications occur efficiently: a training session was organised in less than one week, and further volunteers were encouraged to participate. The rest of the volunteers were recruited through social media and mailshots through University channels. Since most student organisations were actively involved in the festival, we specifically targeted students not belonging to any organisation but that wished to take part in it. A good volunteer recruitment and management strategy is essential for the success of such efforts. Overall there was a very positive response from volunteers: all the 19 recruited volunteers attended the training session held the weeks before the festival and showed up on the night of the festival. During the debriefing with evaluation volunteers at the end of the festival it seemed clear that training sessions were helpful as most of them reported trying some of the approach methods discussed during the training session and adopting the most suitable for them. They also acknowledged the difficulty in collecting randomly sampled data as they would either receive rejections or acceptance from an entire group. An unexpected positive side effect of the training sessions was that although not all volunteers knew each other, throughout the festival a cohesive group formed that encouraged cooperation. The name tag and t-shirt that were handed at the event also helped create a shared identity. During the debriefing at the end of the event some volunteers reported stories of some problems and how they solved them thanks to the others' help.

6.3 One-to-one training for silent observers

The three silent observers were recruited from personal acquaintances of the organisers. They were chosen based on reliability and previous festival experience (as a visitor, as an active participant and never been before). Each one of them could give a different perspective on the festival while trying to be objective. Their diverse backgrounds and experiences helped develop a more complete picture of the festival. Before the festival, the evaluation manager met with each silent observer in a one-to-one session. Their role was explained and discussed and some basic concepts of non-verbal indicators of individual and collective audience reactions were elucidated. The silent observer experience was overall very formative for the festival organisers. Having an external and objective view on the quality of the festival helped analyse mistakes and realise successes better.

6.4 Data Collected

The data collection procedure turned out to be an exhausting experience for the volunteers. This was identified in the debriefing logistics meeting. The volunteers also gave informal feedback that supported the silent observer feedback received. This reason is the most likely explanation for the lower than anticipated number of collected surveys. All interviewers reported being fatigued and inefficient after two hours. This resulted in the final meeting and debriefing session to occur half an hour earlier than planned. This problem could be overcome by splitting the team into two groups and requiring them to collect data for only two hours. This might also improve the quality of the data since visitors would interact with rested and enthusiastic people. By the end of the night, 392 people were directly surveyed through the questionnaire. Information on another 1,237 people was gathered indirectly through group demographic questions administered through the survey (Questions 6 and 7). The data collected was almost three times higher than any previ-

Aim	Survey utility	Silent observer utility
Demographics	Quantitative - Well-sampled data with additional group demographic information gives qualitative insights of the real population.	Qualitative - the majority of the population attending the festival is families.
Reception of the festival	Quantitative - 42% of the people surveyed were going to the festival for the first time. 83% of respondents liked the festival.	Qualitative - impression on the festival and comparison with past editions indicates a decrease in activity quality.
Reason for attending	Quantitative - 76.5% of the sampled people have an interest in science and 59% of them would investigate science further after the festival. Reasons for attending are mainly to have fun (39%), to spend time with the family (26%), and to find out more about University research (21%).	No information.
Interactions	Quantitative - information on interactions of researchers with visitors. While 88% of respondents felt that researchers were enthusiastic, only 54.6% engaged in conversation with them, and 14.6% felt that they were using difficult jargon.	Qualitative - as the night progressed volunteers engaged less with visitors.
Internal feedback	No info on logistics	Qualitative - Organisers received important info on logistics and how/what to improve, for example more clear signage and training stand volunteers to approach audiences.

Table 1: Summary of the multi-strategy evaluation.

ous attempts. Furthermore, notes and comments from the silent observers helped complete the picture of the outcome of the festival. Very constructive feedback was collected for use for future years. For the future of organizing a survey, it is also necessary to collect data regarding how many people were contacted and how many people refused to provide answers. As the data was evaluated, it was evident that the age group between 40–64 prevailed in participating in the responses. Which can lead to skewed overall results. We have no further information as to why this age group formed most of the answers.

6.5 Recommendations for festival organisers

Combining the survey with the information gathered from the silent observers helped gather a good and comprehensive overview of the festival. It set a baseline for future editions on which to build on and keep on improving. All the evaluation efforts carried out in this work were very dependent on volunteers, therefore an effective and appropriate volunteer recruitment, training and retainment strategy was essential. Passive data collection is recommended: any type of data that visitors leave behind naturally such as time spent at a particular location/activity; questions asked; level of engagement through body language; amount of food or drinks sold; number of samples/leaflets handed out; online activity on the website and social media interactions. Once set up, passive data collection is a cheap way of gathering complementary information on an event. Another possibility would be to build-in evaluations in interactive activities carried out throughout the night, such as boards where people can pin their reactions to the event. In SitC, NGOs, government institutions, industry partners and student organisations plan and perform numerous experiments and interactive activities with the guidance of the organisers. With minimal extra effort, a simple and creative way of collecting information on the effectiveness of the experiment or activity could be integrated into the evaluation. Additionally the evaluation team realised the lack of qualitative data. Short informal interviews and/or long focus groups can be conducted with event organisers and visitors to gather impressions and comments to improve future festival editions. A few evaluation volunteers reported that some of the interviewees wished to say more and comment on activities and the festival. Unfortunately there was no time to collect such valuable contributions. The evaluation can also consider the impact on the organising team and external stakeholders: festivals have been proven to help stimulate and maintain partnerships (Bevc et al., 2016). A good understanding of collaborations can help improve and develop them over the years to maximise impact.

7 Conclusions

When considering science communication programmes, practitioners must pay more attention to evaluate their work to assess effectiveness and improve their practices. Ineffective science communication, which is not evidence based, can result in fewer citizens being interested in science or taking a scientific career, which results in the development of greater misconceptions (Gascoigne & Metcalf, 2001; Jensen & Gerber, 2020). Doubtlessly, evaluation of science communication activities are not easy to perform: they depend on the subtle interplay of the audiences, objectives (del Carmen Sánchez-Mora, 2016), communication medium (Cooke et al., 2017; Grand & Sardo, 2017) and interpretations of results which might be misleading (Jensen, 2014). This work has developed a baseline assessment of the festival developing a professional evaluation strategy with limited funds and experience. It is a step-by-step guide for large science festival organisers who want to set up effective evaluation of their efforts. This study used a case study festival to test this approach and managed to identify successes and challenges that helped the organisers improve the festival in subsequent years. Future work will see the iterative evaluation of the festival over a multi-year time span to assess the progression and incorporation of evaluation data by the festival management. Collecting more case studies and tested questions will help improve evaluation for effective implementation.

8 Acknowledgements

The paper's authors thank all the organisations, volunteers, interns and personnel that made this study possible. The Science in the City—European Researchers' Night festival, is organised by the University of Malta and the Malta Chamber of Scientists together with a large number of partners. Funded by the European Commission's Research and Innovation Framework Programme H2020 (2014–2020) by the Marie Skłodowska-Curie actions (grant number: 722824). It is recognized as a Festival by Europe for Festivals and Festivals for Europe (EFFE). It is supported by the Parliamentary Secretariat for Financial Services and Digital Economy and a number of corporate sponsors. The authors would like to thank Ms Silvia Verdolini and Mr Benji Fenech Salerno for providing their expertise and contributions.

9 Privacy and consent statement

9.1 Consent statement

The purpose of the survey was to analyse and present a step-by-step guide on how to set up a valuable and informative mixed method evaluation strategy of large-scale

science festivals and events. The survey was being conducted by the University of Malta. The participation in this survey was voluntary. The participants in this research survey may withdraw at any time. The responses of participants were confidential and we kept it this way for the whole time. When collecting information from participants, consent and permission were explicit, informed, voluntary and re-negotiable. We provided participants with enough information about the data collection that they can make an informed decision about whether they want to participate or not. Everyone had the right to refuse to participate. That meant that if they did not feel like answering the questions, they did not have to, and if they felt like stopping part way through, they could stop at any time.

9.2 Data protection

All data is stored in protected format. To help protect participant confidentiality, the survey didn't contain information that would personally identify a participant. The data are used as a group for statistical purposes and will not be pointed to a single person. We make all data anonymous when analysing it—this means that participants will not be identified.

References

Bevc, C. A., Young, D., & Peterman, K. (2016). Using social network analysis to document science festival partnerships. *JCOM*, *15*(05), A04.

Boynton, P. M., & Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. *BMJ*, *328*(7451), 1312–1315.

Bultitude, K., McDonald, D., & Custead, S. (2011). The rise and rise of science festivals: An international review of organised events to celebrate science. *International Journal of Science Education, Part B: Communication and Public Engagement*, *1*(2), 165–168.

Burns, T. W., O'Connor, D. J., & Stocklmayer, S. M. (2003). Science communication: A contemporary definition. *Public Understanding of Science*, *12*(2), 183–202.

Bussell, H., & Forbes, D. (2002). Understanding the volunteer market: The what, where, who and why of volunteering. *Int. J. Nonprofit Volunt. Sect. Mark*, *7*, 244–257.

Cariveau, D. (2014). Crowd size estimation. *Tratto da course*, *1*.

Cassidy, A. (2006). Evolutionary psychology as public science and boundary work. *Public Understanding of Science*, *15*(2), 175–205.

Cooke, S. J., Gallagher, A. J., Sopinka, N. M., Nguyen, V. M., Skubel, R. A., Hammerschlag, N., & Danylchuk, A. J. (2017). Considerations for effective science communication. *Facets*, *2*, 233–248.

del Carmen Sánchez-Mora, M. (2016). Towards a taxonomy for public communication of science activities. *JCOM*, *15*(2).

Dimitrova, K. (2010). Impact assessment of researcher's night 2010. *Report by ATA48*.

Gascoigne, T., & Metcalf, J. (2001). Report: The evaluation of national programs of science awareness. *Science Communication*, *23*, 66–76.

Grand, A., & Sardo, A. M. (2017). What works in the field? evaluating informal science events. *Frontiers in Communication*, *22*(2).

Hyejin, B., & Ross, S. (2009). Volunteer motivation and satisfaction. *Journal of Venue and Event Management*, *1*(1), 61–77.

Jensen, E. (2014). The problems with science communication evaluation. *Journal of Science Communication*, *1*(C04), 2014.

Jensen, E. (2015). Highlighting the value of impact evaluation: Enhancing informal science learning and public engagement theory and practice. *JCOM*, *14*(3), Y05.

Jensen, E., & Buckley, N. (2011). The role of university student volunteers in festival-based public engagement. *NCCPE*.

Jensen, E., & Buckley, N. (2014). Why people attend science festivals: Interests, motivations and self-reported benefits of public engagement with research. *Public Understanding of Science*, *23*(5), 557–573.

Jensen, E., & Gerber, A. (2020). Evidence-based science communication. *Frontiers in Communication*, *4*, 78.

Jensen, E., & Laurie, C. (2016). *Doing real research: A practical guide to social research SAGE publications*.

King, H., Steiner, K., Hobson, M., & Clipson, H. (2015). Highlighting the value of evidence-based evaluation: Pushing back on demands for 'impact'. *JCOM*, *14*(2), A02.

Koolstra, C. M. (2008). An example of a science communication evaluation study: Discovery07, a Dutch science party. *JCOM*, *7*(2), A03.

Langen, F., & Garcia, B. (2009). Measuring the impacts of large scale cultural events: A literature review. *Liverpool: Impacts*, *8*.

Reja, U., Manfreda, K., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Adv Methodol Stats*, *19*.

Roche, J., Davis, N., O'Boyle, S., Courtney, C., & O'Farrelly, C. (2017). Public perception of European research: An evaluation of European researchers' night in Ireland. *International Journal of Science Education, Part B Communication and Public Engagement*, *7*(4), 371–391.

Sardo, M. (2016). Evaluation of the Bristol Bright Night 2015 (BBN15). *University of the West of England*, *31*, 2017.

Seaton, A. (1997). Unobtrusive observational measures as a qualitative extension of visitor surveys at festivals and events, mass observation revisited. *Journal of Travel Research*, *35*(4), 25–30.

Shaw, I. F., Greene, J. C., & Mark, M. M. (2006). *The SAGE handbook of evaluation SAGE publications Ltd*.

Spicer, S. (2012). *Evaluating your engagement activities. developing an evaluation plan*. The University of Manchester.

Streich, A., Marx, D., Stafford, J., Rodie, S., & Todd, K. (2003). Estimation of attendance at a large outdoor event. *Journal of Extension*, *41*, 5.

Wood, E. (2005). Measuring the economic and social impacts of local authority events. *International Journal of Public Sector Management*, *18*(1), 37–53.

A Appendix – Science in the City, Malta Survey

1. Sex [Female/Male]
2. What is your year of birth? [Year]
3. Where do you live in Malta? Please indicate the letters of your postcode. [Postcode list]
4. Do you have a Maltese nationality? [Yes/No]
5. What is your highest (completed) level of education? [Less than secondary; Secondary; Post secondary; Bachelor; Master; PhD or Professional]
6. Who did you come to the festival with? [Alone; With family; With friends/ Significant other; With colleagues; Other]
7. If not alone, please also indicate the ages and number of people in the group you attended the festival with: [Number of people age 0–15; Number of people age 16–25; Number of people age 26–39; Number of people age 40–64; Number of people above 65]
8. How many year/s have you attended SITC festival before 2016? [Number]
9. How do you rate your general level of interest in science OUTSIDE of the Science in the City festival? [Strongly interested; Interested; Neutral; Not interested; Strongly not interested; No opinion]
10. What are the reasons for attending the Science in the City? [Entertainment; Family time; Fun day out; Learn more about University of Malta research; Meet new people; Other]
11. What is the overall impression of the 2016 Science in the City? [Very good; Good; Neutral; Poor; Very poor; No opinion]
12. To what extent do you agree or disagree to the following statements?

    I felt I was able to participate actively in the Science in the City festival. [Strongly agree; Agree; Neutral; Disagree; Strongly disagree; No opinion]

    I am likely to further investigate some of the scientific topics I encountered at the SITC festival. [Strongly agree; Agree; Neutral; Disagree; Strongly disagree; No opinion]
13. Did the volunteers/researchers at the stands greet you? [Yes/Sometimes/No]
14. Were volunteers/researchers enthusiastic about their research? [Yes/Sometimes/No]
15. Did volunteers/researchers use difficult scientific language? [Yes/Sometimes/No]
16. Did you engage in more (than 5 minutes) conversation with any volunteers/researchers? [Yes/Sometimes/No]

B Appendix – Silent observers instructions

General instructions

- Move around.
- Observe the audience.
- Record information that indicates an event's success (or failure) and the reasons for it.

Keep notes on:

- **Logistics**

  The organisation and scheduling of each event.

  Was it easy to find?

  How are people moving around? Are there points where it is difficult to pass?

  Are people using info points?

- **Audience**

  Make estimates of audience numbers.

  Are people enjoying themselves? Record descriptions of audiences based on non-verbal indicators of individual and collective audience reactions to each event.

  Audience comments: engage in casual conversation, listen to and record spontaneous audience comments.

- **Festival volunteers/researchers**

  Are volunteers/researchers friendly?

  Are they engaging people in conversation? Or are they simply explaining things?

  Do they talk to children? To grown ups? Both?

  Are they organised / neat? Do they give a good impression?

  Were they knowledgeable? Were they understandable?

- **Your opinion**

  Did you enjoy it / find it interesting?

  Suggestions for improvement?

C Appendix - Survey Results

392 survey responses were collected on the night. The statistical relationship between a number of categorical variables were evaluated. Chi-square tests of independence were carried out using IMB SPSS statistics software. For each test, a contingency table was created and the resulting calculations yielded the chi-square value, degree of freedom, p-value and Cramer's V. Where open-ended questions were explored, the qualitative data assessment used was pattern analysis (Jensen & Laurie, 2016) with reliability and objectivity addressed through independent coding.

Interactions with researchers/volunteers

Greetings

Greetings	**Number of responses**	**Percentage**
Yes	313	79.8%
Sometimes	35	8.9%
No	27	6.9%
N/A	17	4.3%

Jargon

Jargon	**Number of responses**	**Percentage**
No	245	62.5%
Sometimes	68	17.3%
Yes	57	14.5%
N/a	22	5.6%

Enthusiasm

Enthusiasm	**Number of responses**	**Percentage**
Yes	346	88.3%
Sometimes	28	7.1%
No	2	0.5%
N/A	16	4.1%

Conversation

Conversation	**Number of responses**	**Percentage**
Yes	214	54.6%
No	104	26.5%
Sometimes	54	13.8%
N/A	20	5.1%

Table 2

Demographics

Total number of people surveyed 392
Female 48.7%; Male 51.3%

Age (by category)

Age	Number of people	Percentage
0–15	88	22.4%
16–25	95	24.2%
26–39	48	12.2%
40-64	118	30.2%
65+	43	11.0%

Min = 6 years old; Max 96 years old

Education

Highest Education	Number of people	Percentage
Post Secondary	102	26.0%
Bachelor	93	23.7%
Secondary	69	17.6%
Less Secondary	85	21.7%
Master	30	7.7%
PhD/Professional	11	2.8%
NA	2	0.5%

Where they live in Malta

Place	Number of people	Percentages
"ATD"	35	8.9%
"BKR"	24	6.1%
"NXR"	21	5.4%
"SGN"	21	5.4%
"NA"	16	4.1%
"MST"	16	4.1%
"SLM"	12	3.1%
"ZBG"	12	3.1%
"SPB"	11	2.8%
"Not Malta"	11	2.8%
"ZRQ"	10	2.6%
"BBG"	9	2.3%
"FGR"	9	2.3%
"MSD"	9	2.3%
"ZBR"	9	2.3%
"HMR"	8	2.0%

Disclaimer: Postal addresses below a frequency of 8 are not included.

Table 3

Group demographics

Age of accompanying people (by category)

Age range	Number of people	Percentage
0-15	436	35.2%
16-25	355	28.7%
26-39	195	15.8%
40-64	220	17.8%
65+	31	2.5%

Total number of people: 1237

Estimate total number of people by category, which age category is the largest population at the festival?

Age range	Number of people	Percentage
0-15	524	32.2%
16-25	450	27.6%
26-39	243	14.9%
40-64	338	20.7%
65+	74	4.5%

Total number of people: 1629 (direct + indirect)

Type of accompanying people – Who did you come to the festival with?

Type	Number of responses	Percentage
With family	224	57.1%
With friends/ Significant other	134	34.2%
Alone	20	5.1%
With Colleagues	9	2.3%
Other	3	0.8%
N/A	2	0.6%

Table 4

Reception of the festival

Returning audience [% of the people that has been there before; % that came more than once]

Frequency	Number of people	Percentage
0	162	41.3%
1	85	21.7%
2	70	17.9%
3	44	11.2%
4	14	3.6%
5	5	1.3%
6	7	1.8%
N/A	5	1.3%

Did they like it?

Impression	Number of people	Percentage
Good (G)	168	42.9%
Very good (VG)	157	40.1%
Neutral	44	11.2%
Poor (P)	7	1.8%
No opinion	5	1.3%
Very poor (VP)	3	0.8%
NA	8	2.0%
VG + G = 83%		
VP + P = 2.7%		

Reason for attending

Reason	Number of people	Percentage
Family time	102	26.0%
Research in Malta	85	21.7%
Entertainment	84	21.4%
Fun day out	69	17.6%
Other – learn more	13	3.3%
Other – by chance	10	2.6%
Other	7	1.8%
Other – curiosity	5	1.3%
Other – participation	4	1.0%
Meet new people	3	0.8%
Other – particular exhibit	2	0.5%
N/A	8	2.1%

Interest in science

Level of interest	Number of people	Percentage
Interested	162	41.3%
Strongly interested	138	35.2%
Neutral	62	15.6%
Not interested	16	4.1%
Strongly not interested	4	1%
No opinion	2	0.5%
N/A	9	2.3%
SI + I = 76.5%		
SNI + NI = 5.1%		

Follow up on science

Opinion	Number of respondents	Percentage
Agree (A)	159	40.6%
Strongly agree (SA)	72	18.4%
Neutral	80	20.6%
Disagree (D)	48	12.2%
No opinion	12	3.1%
Strongly Disagree (SD)	7	1.8%
N/A	13	3.3%
SA + A = 59%		
D + SD = 14%		

Table 5