Fighting the Machine: Co-constructing Team Based Evaluation for Non-Formal Learning

Andrew Clapham, Raquel Barata***

Abstract

Museum educators increasingly face demands to evaluate the 'value' of non-formal learning (NF-L). This paper offers a unique international and multi-educator perspective on how informants from Portugal, Italy and the United Kingdom navigated these demands. Analysis of interview data highlighted that, although working in three different countries, most of these educators had experienced evaluation as accountability (and disciplinary) focused, employing methodologies inappropriate for evaluating NF-L and rarely team based. Drawing on a composite theoretical framework, these data led to co-constructing the Team Based Evaluation (TBE) model. Two case studies map how TBE was enacted and recommendations concerning organizational change are made. The paper concludes that, whilst set within the museum education space, TBE can be applied across evaluation contexts and micro and macro scales. Locally, TBE can mediate rich evidence and develop teamworking practices. Nationally and internationally, it can contribute to resetting evaluation from an accountability and disciplinary 'machine' to a dialogic, democratic and developmental activity.

Keywords: Non-formal Learning; evaluation; performativity; accountability; dialogic; democratic; development.

Introduction

Almost 25 years ago, John Falk and Lynn Dierking (2000) asked the question 'Why do people go to museums and what do they learn there?' In this paper, we explore this question via the policies, processes and practices of museum educators who evaluate non-formal learning (NF-L).

Internationally, cultural education settings such as museums, botanic gardens, science centres, theatres and galleries face demands to evaluate their 'value for money'. Whereas many are proactive in evaluating exhibitions and projects, they are less likely to evaluate NF-L (Clapham 2023). Whilst this is partly down to a lack of resourcing and expertise (Kubarek 2015), NF-L's complexity only adds to the challenges for those evaluating it (Clapham 2023).

The paper offers a unique international and multi-educator perspective on how museum educators from Portugal, Italy and the United Kingdom undertook NF-L evaluation. Analysis of interview data mapped the landscape within which NF-L evaluation occurred in these counties and underpinned the Team Based Evaluation (TBE) model. Although working in three different countries, informants reported that evaluation was regarded as threatening and disciplinary (see Preskill 2011); rarely dialogic, democratic or developmental, (see Bulaitis 2020) and seldom if ever team based (see Clapham 2023).

Many of the informants' discourses revolved around how evaluation was to be feared rather than embraced – either because of its use as a performance management tool, or because they lacked the skills and confidence to undertake it. Unfortunately, for many, the way evaluation was employed in their organizations resonated with Kubarek and Trainer's (2015: 3) description of an 'outcomes-based work environment… riddled with accountability

and pressure'.

TBE was developed to counter these deficit models and can be employed by educators, education departments – and museums more widely – to develop their evaluation process and strategy. Two case studies highlight how educators, both heads of museum education departments, enacted TBE and we offer recommendations – based upon organizational maturity theory (Crosby 1979; Clapham 2024) – as to how the organizational change required to enact TBE can be achieved.

We argue that, although set within the policy and accountability space of museum education, TBE is applicable across scales and contexts. Locally, it can produce rich evidence and develop teamworking practices which have positive benefits beyond NF-L evaluation. Nationally and internationally, TBE can contribute to resetting evaluation from an accountability and disciplinary 'machine' (Dahler-Larsen 2011) to a dialogic, democratic and developmental activity.

Policy context

Evaluating museums' impact and 'value for money' is increasingly high stakes. For example, in 2022 the United Kingdom's (UK) Department for Education published the *Opportunity for All* White Paper, which included the *National Plan for Cultural Education* -– the first UK wide cultural education policy for ten years. *Opportunity for All* stressed the need to evaluate the 'economic benefits of the white paper's ambitions', which exemplifies requirements to evaluate the economic value of cultural education not only in the UK but also internationally (see Kubarek and Trainer 2015; Clapham 2023).

What narratives such as *Opportunity for All* seem to ignore is the complexity of defining NF-L (see Johnson and Majewska 2022) and thus the complexity of evaluating it (Barata *et al*. 2017). Nonetheless, 'performative' (Lyotard 1984) demands to demonstrate the value of learning and education – exacerbated by economic uncertainty (Curioni and Rizzi 2016) and the global COVID pandemic – increasingly drive why evaluation is undertaken (Dahler-Larsen 2011; Clapham 2023). Educators are therefore caught between competing tensions around why and how they evaluate – they must 'prove' NF-L's worth in performative terms, whilst also acknowledging that its complexity raises significant questions as to why (and how) one would do so.

Theoretical framework

In this section, the project's composite theoretical framework is discussed. Performativity was used to understand the wider evaluation policy context and Team Based Inquiry (TBI) and Thinking Evaluatively (TE) underpinned the TBE model.

Performativity

Performativity is the optimization of efficiency through inspection and ranking (Lyotard 1984) and has been widely used as a concept for understanding education systems (see for example Ball 2003; Clapham *et al*. 2016). Although most of the work around performativity and education has been focused on 'formal' settings such as schools and universities, museums and other cultural education settings have increasingly faced performative pressures (Kubarek and Trainer 2015; Clapham 2023).

Central to Lyotard's argument is that learning settings such as museums are not the same as industrial settings such as factories. Consequently, even though performativity might make sense on an industrial production line, it becomes highly problematic when applied to education systems (see also Ball 2003). Lyotard contended that performativity effects how education systems are organized and shapes what educators do – what Ball (2003) describes as the 'terrors of performativity'.

These performative 'terrors' are so powerful that they have profound outcomes for educators' 'inner-life' (Ball 2003: 226) and their professional selves. This has been brought into focus in the UK after the suicide of a headteacher (Ruth Perry), which the case coroner found to have been a direct result of the grading given to her school by the national school

inspectorate (the Office for Standards in Education, Children's Services and Skills (Ofsted)).

Lyotard's work paints a vivid picture as to how performativity operates as a highly powerful mechanism for organizing how society operates, what people do and what knowledge is considered most 'legitimate'. Whilst writing in the 1970s, his analysis appears to have been borne out in contemporary practice, with performative technologies – and their use of comparisons, judgements and rankings – accepted as the 'norm' in almost every facet of modern society.

Team Based Inquiry (TBI)

TBI was initially developed to support informal science educators to 'develop the skills and tools to incorporate evaluation and data-informed decision-making into their work' (Cohn, 2019, n.p). TBI involves an 'evaluative cycle' of asking questions, undertaking investigation, reflecting on findings, and improving practice that educators can realistically undertake. This cycle is (i) systematic; (ii) led by non-evaluation professionals; (iii) collaborative and team-based; and (iv) small-scale and short-term (see Cohn, 2019).

Key to TBI is that it aims to be a practical means for embedding evaluation in educators' work so that they can

> '…get the data they need, when they need it, to improve their products and practices and, ultimately, more effectively engage public and professional audiences' (Pattison *et al.* 2013: 5)

As much as TBI is a practical approach toward evaluation, it also has additional benefits for education departments and museums more widely:

> Over and over, we have found that a collaborative process of inquiry helps team members achieve shared understandings of their educational goals, discuss underlying assumptions and expectations, find effective strategies for working together and achieving consensus, and develop trust and a sense of common purpose (Pattison *et al.* 2013: 6-7)

Cohn (2019) also describes TBI as a 'truly collaborative process', which enables educators, education departments and museums to accomplish organizational goals and objectives. The collaborative aspect of TBI is particularly powerful in that it values differing perspectives; enables team members to learn from each other; fosters improved efficiency; cultivates communication and strong working relationships; and leads to an expanded sense of accomplishment (see Pattison *et al.* 2013).

Despite these positive outcomes, designing organizational processes and practices so that stakeholders work together is challenging. Nonetheless, research evidence exploring evaluations undertaken in teams (for example Adams 2012) suggests that the results can be transformational both for how evaluation is undertaken and for working practices across education departments and entire museums.

Thinking Evaluatively (TE)

TE can drive evaluation strategy at the scale of the Education Department and the museum and is:

> '…motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and informing decisions in preparation for action' (Buckley *et al*. 2015: 378).

The 'context of evaluation' not only includes formal evaluations but also 'informal evaluative efforts that inform and improve actions' that take place 'throughout all of an organization's functions' (Buckley *et al*. 2015: 378). Kubarek (2015: 9) similarly argues that TE can 'foster an environment of collaboration through transparent practices' which in turn can build capacity so staff can 'take an active role and ownership of the evaluation process'.

TE advocates that effective evaluation involves a wide range of team members, who understand what it is to be an evaluative thinker (Buckley *et al*. 2015). TE argues that evaluation is recognized (across a museum) as a complex endeavour that requires commitment and capacity building (Kubarek and Trainer 2015). Unfortunately, as Kubarek (2015) argues, this is often not the case; more frequently, educators face a lack of time, resources, or experience that, when combined, can make embedding TE highly problematic.

Of course – and as Buckley *et al*. (2015) outline – the absence of TE does not mean that evaluation ceases to occur. However, such projects run the risk of asking poor questions, generating and analysing inappropriate data and ignoring unexpected developments (see Dahler-Larsen 2011). Additionally – and as Dahler-Larsen (2011) continues – evaluations which are designed without a sense of evaluative thinking can stagnate due to a lack of motivation of those involved; implementation of recommendations can fail to take place; and often there is resistance to change.

Clearly, whilst TBI and TE do not offer a 'golden bullet' for addressing these challenges, they do highlight the power of team based evaluation that is underpinned by an evaluative thinking mindset across a museum. As such, both TBI and TE resonate strongly with Garcia's call for museums to revisit their 'learning power':

> …when museums describe their educational impact to stakeholders, it is often described narrowly, using the measures of formal education rather than focusing on its capacity to model intrinsically-motivated, joyful, open-ended learning… Museum educators are not doing enough to make a case for the value of museum learning in its own right... (Garcia 2012: 47).

Drawing on Garcia's (2012) comments – and the work of Kubarek and Trainer (2015); Kubarek (2015); Dahler-Larsen (2011); and Buckley et al. (2015) – we outline how team based evaluation can reset 'narrow' accountability focused descriptions of NF-L. Although not a quick-fix, both TBI and TE can support educators to develop rich evaluations in the NF-L space as well as empowering educators to navigate the demands of performative evaluation.

Methodology

The project took place across two phases. Informants for both phases (Table 1) were part of a larger cohort (*n*=80) of educators from Italy, Portugal and the UK participating in the Learn to Engage (LtE) project. In the first part of Phase I, semi-structured reflexive interviews (*n*=12) with educators (*n*=12) described the policy environment within which NF-L evaluation took place. The second part of Phase I saw analysis of these data (along with the theoretical underpinning afforded by TE and TBI), which led to the TBE model. In Phase II, two 'case studies' (Yin 2003) were undertaken with educators – both heads of education departments – to explore how they enacted TBE in their museums.

Phase 1		
Experience	**Number of informants (n=80).** NB: of these, data were used from 12 informants.	**Number of interviews (*n*=47).** NB: of these, 12 interviews were used.
Head of museum education department	(*n*=24, 30%)	9
Educators with more than 10 years' experience	(*n*= 36, 45%)	15

Educators with between 2 and 10 years' experience	(*n*=8, 10%)	17
Educators with less than 2 years' experience	(*n*=8, 10%)	4
PhD students and senior leaders	(*n*=4, 5%)	2
Phase 2	**Julia**	**Richard**
Informant Biographies	Julia is a museum professional, in her early fifties, responsible for activities at her education department for 24 years. She participated because the museum had not implemented evaluation strategies.	Richard is a botanic garden professional in his fifties, botanist and science educator for 25 years and Head of education for 15 years. He holds academic qualifications and has major experience in the role. He decided to participate in the research since his team already implemented some evaluation tools and methodologies.

Table 1. Informants

Phase I interviews were captured via reflexive field notes (see Kvale and Brinkmann 2009), whilst Phase II interviews were captured via reflexive field notes and recordings via MS Teams. Interviews 1 and 2 with Julia (the names 'Julia' and 'Richard' are pseudonyms to protect the informants' identity) took place 'face-to-face' in May 2021 and July 2021, with interview 3 in June 2022. Interviews 1-5 with Richard took place between May 2021 and October 2022, with interviews 6 and 7 in November 2022. Data were analysed via reflexive thematic analysis (Braun and Clarke 2021) (see Table 2).

Phase I Themes	**Evaluation Domain**	**Challenges**	**Gap**
Working as a team	Define the evaluation Undertake the process	Resources Motivation Purpose of evaluating Accountability versus development	PD and training
Resourcing	Undertake the process	Practicalities Requirement for impact and hard data Acknowledging educators' expertise Motivation	Resourcing

Voice and trust	Define the evaluation Undertake the process	Valuing educators and visitors as evaluation co-constructors as well as informants Being prepared to hear and valuing what informants say	PD and training
Communicating and implementing findings	Act, interpret and reflect on findings	Lack of diverse evaluation audiences Wrong questions, evaluated in the wrong way Lack of change	Expertise and team engagement
Organisational buy-in	Define the evaluation Undertake the process Act, interpret and reflect on findings	Type of evidence required for buy-in Making causal claims Methods and analysis Forced pragmatism Playing the game	

Table 2. Understanding the Evaluation Space: Thematic Analysis

In the following section, the themes resulting from analysis of Phase I informants' data are presented.

Understanding the evaluation space

The '*Working as a team*' theme reflected the concerns of informants from all three countries about the lack of opportunities to undertake team based evaluation. However – and again across all three settings – this theme was also indicative of challenges in engaging in teamwork more generally:

> We've got such a small team now… doing any work together that's not directly to do with visitors… is almost impossible. We don't have time to plan sessions together, let alone evaluate them together (Informant *G*).

Other informants also reported how their departments had significantly reduced in size, which reduced capacity to undertake evaluations and the amount of evaluative expertise held by department members. This erosion of evaluative knowhow was further compounded by a lack of opportunities for professional development (PD) around NF-L evaluation:

> …PD opportunities [concerning evaluation] … they're extremely rare or non-existent. And when they are available, they don't address the complexities of evaluating NF-L (Informant *Q*).

Most informants also reported an absence – or significant gaps – in departmental or museum wide evaluation strategies. When these strategies were in place, they positioned evaluation primarily as a means for evidencing accountability:

> We do have evaluation polices, but they're about performance management… they're not about developing NF-L (Informant *H*).

Informants' concerns about this lack of opportunities for teamworking linked to the '*Resourcing*' theme. Again, educators working across Portugal, Italy and the UK reported similar climates of reduced funding and increased requirements to evidence economic impact. A recurring

point was how this climate affected almost every facet of their department's work, not just NF-L evaluation. Many recounted how reduced funding meant that resourcing team based activities of any sort was progressively more problematic:

> We've had so much funding cut…. Just doing our core work is difficult enough. We simply can't work as a team because there aren't enough of us to do it (Informant *B*).

Informants' unhappiness about the lack of teamworking opportunities was further intensified in that their museums did resource (performance management focussed) evaluations undertaken by external consultants:

> I'd love to be able to resource my department to work together [to undertake TBI]. My line manager tells me that there's no money for the Education Department though… and the next thing I know, I'm doing a questionnaire from a consultant for an impact evaluation! (Informant *F*).

Informants made strong associations between evaluation and accountability, which was also reflected in the '*Voice and Trust*' theme. For many educators, the link between evaluation and accountability had become embedded within the practices and processes of their organizations (be it located in Portugal, Italy or the UK). One outcome of this was that their voices were not heard and their expertise (as educators) was not trusted as part of evaluation projects. Informants also felt that the voices of visitors – and crucially non-visitors (see Kluge-Pinsker and Stauffer 2021) – were also disregarded:

> It's all about numbers and targets and value for money. They [evaluation audiences] don't want to hear our voices or the visitors'… (Informant *J*).

Informants also described how they felt their contribution to projects as evaluators appeared to be distrusted. Although most acknowledged that they would like to be more skilled as evaluators, they were frustrated that their expertise as education professionals rarely featured in evaluations. For many informants, this was a direct result of evaluation being used for accountability:

> We don't feel like we're trusted [by the museum management]… because evaluation is all about our performance, they don't want us to "mark our own homework". I don't mind having an annual review, but I get frustrated when evaluation stops being about developing learning (Informant *S*).

This lack of trust also appeared in the '*Communicating and implementing findings*' theme. Again, across all three countries, informants reported a two-tier system around implementation. Findings from 'top-tier' evaluations (that focused on economic impact and had policymakers, funders and government for their audiences) were highly likely to be implemented. In contrast, findings from 'bottom-tier' evaluations (that focused on developing pedagogy and had other educators or learners for their audiences) were far less likely to be implemented:

> We only seem to report [evaluation findings] to funders and tell them what they want to hear. What we've stopped doing is reporting to the visitors (Informant *C*).

Most informants felt that this disparity was due to a lack of communication and implementation strategies being clearly articulated at the outset of evaluation projects. Informants reported that as a result, evaluations were considered to be complete (by both evaluators and management) with the publication of the final report, rather than when recommendations had been implemented:

> The whole point of evaluation is that it's about development… that it leads to change… findings need to be implemented… and there's a will to accept change. If not, then it's been a waste of time (Informant *F*).

The final theme concerned '*Organizational buy-in*'. Here, informants outlined how museum education in the three countries in which they worked was being side-lined and that education

departments had borne the brunt of many funding cuts. This lack of buy-in reflected far more than just omissions in their own museums' evaluation strategies:

> The museum [leadership] considers learning way down its list of priorities… which is crazy, because museums are about learning! Museum education isn't taken seriously (Informant *K*).

Many informants also felt that the absence of buy-in for team based NF-L evaluation was indicative of how museum education was being diminished – a view held by educators in all three countries:

> I really fear for museum education… I think that the more we try to evaluate NF-L through numbers and impact the less we're able to really understand it. And if we can't understand it [NF-L] … well eventually that will mean we don't understand why we have museums at all (Informant *W*).

Informant W's comments were typical of how many others saw the future of museum education not only in Portugal, Italy and the UK but globally. They felt there were clear links between requirements to demonstrate the (economic) impact of museums and the diminishing of museum education, and were therefore indicative of the pressures facing the entire cultural education sector.

Team Based Evaluation in Practice

Table 3 indicates the main themes that emerged from the Phase I interviews and which underpinned the co-construction of the TBE (Figure 1). TBE is based on three 'evaluation domains' that highlight how it is a group-mediated, organic cycle embedded within daily practice. Figure 1 illustrates how TBE provides structure to NF-L evaluation, whilst also signalling potential challenges and gaps within departmental and organizational evaluation processes.

Figure 1 also reflects how the term 'evaluator' should become defined in a far more holistic manner. Currently, 'evaluator' appears to mean 'expert from outside an organization'. TBE contends that an evaluator can be an external expert such as an evaluation consultant; however, TBE groups evaluating NF-L should consist of a wide membership including: learners; parents, volunteers, museums members; teachers; policy makers; senior leaders; marketing departments; visitor development, exhibition and interpretation design teams; and curators (this is not an exhaustive list).

Categories	Themes	Data
TBE is active	Team working Doing TBE Implementing findings and making changes Buying in	Educators feel valued as contributors to the evaluation design, data generation and analysis stages, the development of findings and empowered to implement recommendations.
TBE is autonomous	Team working Doing TBE	TBE is a group activity with autonomy and independence from internal and external drivers that might want to influence findings and recommendations.

TBE values multiple voices	Team working Buying in	TBE values the voices from a wide variety of team members from within and outside the museum including: marketing departments, visitor development, exhibition and interpretation design, curators and the membership, learners, parents, teachers, visitors and non-visitors as well as consultants, policy makers, funders.
TBE cares	Doing TBE Buying in	TBE members are invested in thinking evaluatively. They highly value the degree to which evaluators care about what is being evaluated.
TBE does not offer a quick-fix	Implementing findings and making changes Buying in	TBE offers a strategic framework for the mid and long term development of NF-L activities across a museum.
TBE is legitimate	Implementing findings and making changes Buying in	TBE is regarded by the museum leadership as at least equal to other evaluation [illegible] dership understands the [illegible]NF-L and considers TBE [illegible]nd rich means of evaluating [illegible].
TBE heightens awareness of findings and	Team working Doing TBE Implementing findings and making changes Buying in	TBE leads to a wider understanding of why evaluation is taking place, what findings the evidence leads to and what change might be required to implement recommendations.

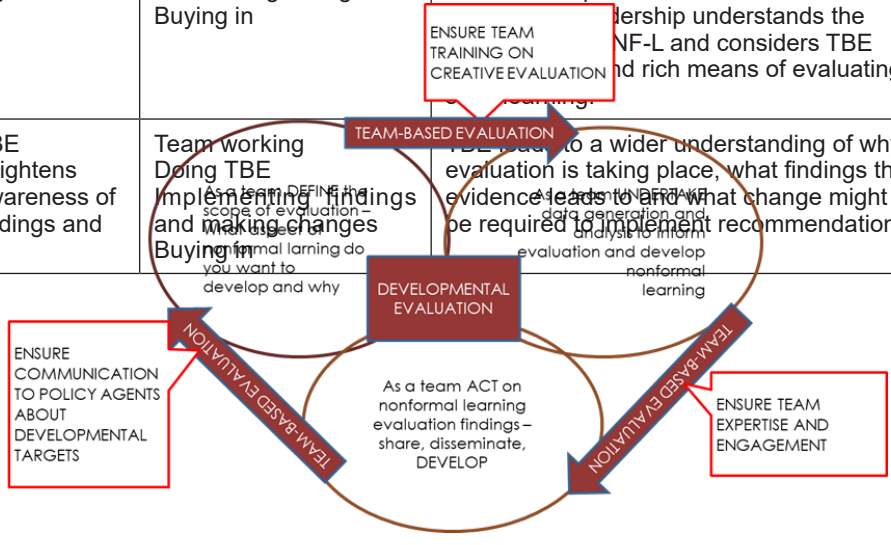

Table 3. Enacting TBE: Thematic Analysis

Figure 1. Modeling TBE

Example of intervention being evaluated	**Example TBE question**	**Example TBE methods**	**Example TBE Audience(s)** (can also be TBE group members**)**
Science cafes	What scientific concepts have been learned during the science cafe?	Visitors invited to be part of science café project team and TBE group members. Science café design, focus and evaluation strategy co-constructed with TBE group (including with audience). Development of participatory tools to collect data. Key evaluation landmarks identified over science café duration – data generated and analysed, interim findings made and disseminated. Feedback sessions built into programme along with dissemination of findings by TBE group (including visitors) to museum leadership at prearranged points.	Adult visitors to science cafes. Not experts, but interested in the scientific themes being approached

School group activities	To what extent are teachers satisfied with the activity and what suggestions to improve?	Teachers, learners and parents invited to be part of project team and TBE group from outset. TBE group (including teachers and learners) co-construct the focus of the activities as well as the evaluation strategy. Development of multi-modal creative methodologies: dance; poems; installations; teacher/ learner video diaries. Key evaluation landmarks identified during the activities – data generated and analysed, interim findings made and disseminated Feedback sessions built into programme along with dissemination of findings by TBE group (including teachers, learners, parents) to museum and school leadership (including governing body) at prearranged points.	Teachers, learners, headteachers, parents, school governing boards

Table 4. TBE Example Projects

In these next sections, data from the Phase II interviews with Julia and Richard are discussed (see Table 4 for example TBE projects).

Teamworking

Both Julia and Richard stressed that TBE could facilitate rich NF-L evaluations in ways difficult to achieve when working alone. Moreover, TBE groups could foster positive outcomes from teamworking that went beyond evaluating NF-L:

> Working as a team [to evaluate] helps to develop links between the team members… it's rewarding for the team to accomplish common goals… only teamworking can highlight common goals (Richard).

These links resulted from TBE group members valuing each other's expertise, not just as museum educators but from across the museum (such as curators or exhibition designers) and external to the museum (learners, parents, teachers). This teamworking also went some way to alleviating the often threatening and disciplinary ways informants experienced evaluation (see also Preskill 2011). Consequently, TBE group members were less cautious, stressed and demotivated by evaluation; more invested in evaluation findings; and more likely to implement recommendations and embrace change.

TBE groups also enabled educators to develop their confidence as evaluators. Both informants felt that as a result, evaluations were stronger in both design and delivery and there were greater levels of motivation to evaluate. Perhaps most crucially, it was more likely that findings would be implemented with less resistance to change.

Whilst highlighting its benefits, both informants were pragmatic about the challenges in enacting TBE. Most pressing of these were the practicalities of resourcing teamwork, not just for TBE groups but across their department's work. Julia outlined how the size of her

team (three full time and two part time members) meant that teamworking was challenging even in day-to-day 'core' work:

> …we could perform evaluation on the NF-L about a specific subject [for example, deforestation], but we don't have teams big enough.

Most of Julia and Richard's comments around TBE groups, and teamworking generally, reflected their wider concerns around the future of museum education *per se*. Both highlighted the need for increased resources to support teamwork not only as NF-L evaluators, but for core academic, pedagogical and vocational activities. Similar to Phase I informants, both regarded these concerns as indicative of an existential threat toward museum education.

Doing TBE

'Doing' TBE enabled group members to engage in a reflexive process not only around NF-L evaluation, but also museum education more widely. As a result, educators were able to develop clearly defined aims and objectives for evaluations, which led to group members sharing the same values (around data and methodologies for example). Working this way also meant that a framework for the evaluation was agreed by group members, who both valued and abided by it.

This strategic and shared approach toward developing and undertaking evaluations meant that direct connections between these projects and wider departmental and museum goals were apparent. This shared evaluation framework also meant that the norms around why and how evaluation took place were developed and valued. For example, evaluation developed via TBE groups was not a top-down initiative used to evidence performance management for individual educators. As a result, TBE projects produced rich evaluative evidence and (just as importantly) helped to re-establish trust in evaluations.

Once again, both Julia and Richard considered that 'doing' TBE reflected far larger debates concerning museum education than solely evaluating NF-L. Both described how locally, nationally and internationally the value of museum education was increasingly focused on its economic impact. Museum educators therefore had to negotiate demands for demonstrating impact, whilst maintaining their beliefs that museum education was core to their museum's mission:

> 'High policy agents' want numbers to justify quality, but usually the Direction Board want more information… they're only interested in numbers… (Julia).

> …evaluation has to provide the hard numbers that [the] Director needs to argue the case for impact (Richard).

Both informants felt that 'doing' TBE required a highly sophisticated and strategic approach, as evaluations had to negotiate demands to 'play the numbers game' (Richard), whilst remaining a means of developing NF-L.

Implementing findings and making change

Both informants felt that TBE could prevent inequity in how evaluation findings and recommendations were (or were not) implemented. They described historical failure to implement evaluation findings that had been undertaken by their departments and that were focussed on developing NF-L pedagogy. In contrast, evaluations focused on accountability, impact or improving efficiency had been implemented.

To counter this lack of equity, TBE groups can develop strategic approaches toward implementing evaluation findings and leading change from the outset. This was achieved through clear dialogue between TBE group members and other stakeholders (including the museum leadership) throughout an evaluation project, not just at its conclusion. TBE groups can also develop and communicate clearly defined short-term and long-term goals, realistic timelines and tangible steps toward implementing findings.

Both informants stressed the importance of this communication and that communication

strategy should be inbuilt throughout an evaluation. Whilst ostensibly such communication might appear unproblematic, both described how it was all too easy for evaluation teams to underestimate or simply miss its importance. Both informants were clear on the crucial importance of communication in TBE – evaluation findings were far more likely to be implemented (and less resistance to change encountered) when evaluation was undertaken via TBE groups with clear communication strategies.

This notion of resistance to change was an important point. TBE stresses that clear, open and honest communication between TBE group members and other stakeholders supports how the organizational change resulting from evaluation is understood. In part, this is due to TBE groups co-constructing evaluations, which meant that changes resulting from recommendations ceased to be top-down, externally imposed or unexpected. TBE group members therefore drive the implementation of findings and have an 'inside view' as to why change was taking place. This was important in countering fear of organizational change which is often strongly associated with negative events such as reduction in funding and redundancies (see Wissema 2000).

Buying-in

Both informants reported challenges around attaining organizational and stakeholder buy-in for TBE. Whilst this could be ascribed to wider challenges around funding, they felt that the entire concept of TBE appeared problematic for some stakeholders. The main reason for this was that unless evaluation had a clear focus upon impact, buy-in for other modes of evaluation was unlikely:

> The museum is not a priority for decision makers, but they are interested in numbers… [but] they don't have any background in museums (Julia).

Both informants described how, because TBE was not the 'norm', there was scepticism – from their museum's leadership – as to its effectiveness. Their museums considered that the 'safe option' was to employ external consultants, often working in isolation, to undertake impact evaluations. Although there were no 'quick fixes' for attaining buy-in, both felt that there were strategies that can support the organizational change required for TBE buy-in to be successfully gained (later we discuss Organizational Maturity Theory (Crosby 1979) and Governance Maturity Theory (Clapham 2024) as two examples).

Despite the challenges in attaining it, both informants felt that when stakeholder buy-in was in place, its benefits operated across several levels. Most obviously, buy-in meant that resourcing TBE groups and evaluation projects was less problematic. However, buy-in also had more nuanced outcomes; for example, possible tensions between stakeholders resulting from unexpected or unpalatable evaluation findings and recommendations were less likely to occur.

Discussion

Informants' data highlight how TBE and TBE groups can undertake informative NF-L evaluations that can lead to positive change. Group membership can positively impact upon educators' experience of evaluation and potentially across other aspects of education departments' work and museums more widely. TBE groups – underpinned by a departmental/museum wide evaluative mindset – values diverse expertise and voices and considers collaboration between a range of informants and voices (and input from 'expert others' from outside the museum) as fundamental to evaluating NF-L. Perhaps most significantly, TBE and TBE group membership can also increase levels of ownership (and trust) of NF-L evaluations which transforms them from a form of accountability to a developmental activity.

Despite these positives, enacting TBE can face significant obstacles, with educators' lack of confidence as evaluators a prominent example. This lack of confidence is particularly relevant as NF-L evaluations are often undertaken alone, by educators relying on their own evaluative expertise. Consequently, educators felt they do not possess the skills and confidence to be able to evaluate NF-L effectively, which meant they were demotivated and stressed

when having to do so. A lack of continuing professional development opportunities (which if available would have only catered for a generalist view of evaluation rather than being NF-L focused) only added to this problem.

A lack of organizational buy-in for TBE was also a significant obstacle. Although TBE can be enacted by educators and education departments on their own, this is significantly more complex than if organizational buy-in is in place. This lack of buy-in signalled for many informants far more than simply a reluctance to instigate a new mode of evaluation. In the cases reported here, informants argued that education was considered secondary to other museum activities which presented significant obstacles toward gaining buy-in for TBE.

Educators attempting to enact TBE also faced obstacles from what they felt was a narrow view as to who should be the audiences of evaluation reports. Unsurprisingly perhaps, these audiences are 'major players' such as trustees, funders and policy makers. However, this focus contrasts with that of TBE, which considers learners to be at least equally significant.

The final obstacle was overcoming educators' scepticism as to the purpose of evaluation. This mostly resulted from the lack of previous evaluation findings (most notably related to pedagogy) being implemented, whilst impact evaluations had been. When this had occurred, some educators considered TBE as just another performative tool that fails to lead to change.

Evidently, there are 'pros and cons' related to enacting TBE as policy, process and practice. On the 'pro' side, TBE and TBE groups were considered powerful approaches toward undertaking NF-L evaluation. On the 'con' side, there were significant obstacles that had to be overcome for TBE to gain traction. In many ways, these obstacles are as much related to organizational culture – and the wider climate facing museums – as they are to TBE as an evaluation strategy.

Achieving organizational change to successfully enact TBE can be problematic; however, there are numerous change models that can help mediate this. For example, Organizational Maturity Theory (OM) (Crosby 1979) has been successfully used to mediate organizational change in a range of settings. Similarly, Governance Maturity Theory (Clapham 2024) has been used to drive change in school governing bodies. Table 5 highlights the key drivers mediating successful change and how these can be applied to enacting TBE in museums.

Change driver	**Organizational change for TBE enactment**
Accepting the need for change	Must be acceptance that change is required, which in turn means there is commitment to change. This means that not only TBE advocates but also those across the museum must see the value in – and be willing to fully commit to – new ways of working, behaving, and thinking.
Clarity of vision for change	Changes required for TBE to be used must be clearly and simply stated so that all stakeholders – including educators and museum leadership – can see, understand, and embrace the big picture around the importance of evaluating NF-L, what is involved and why TBE is imperative.
Consistent and compelling messaging	There must be consistent and compelling messaging around why TBE is important and what it offers the entire museum, not just the Education Department. Doing so means not only that educators are TBE advocates but its value is seen across the museum.
Modelling changed behaviours	Organizational change can be achieved via the modelling of desired TBE behaviours. Doing so encourages the behaviours that lead successfully to embedding TBE through the museum. This can be achieved by role-modelling TBE behaviours and recognizing and rewarding those who embed TBE within their evaluation strategies.

Staying on track	Role-modelling TBE behaviours and recognizing and rewarding those who embed TBE within their evaluation strategies prevents either stagnation of change or slipping back into previous evaluation routines. Ultimately, staying on track means that the 'new' evaluative behaviours inherent in TBE become the 'normal' behaviours.

Table 5. Key Drivers for Organizational Change

OMT argues that maturity 'stages' – on a continuum from immaturity to maturity (Table 6) – can be used to map and develop an organization's processes and practices over time. These stages are not destinations but progressive steps and even highly successful organizations are not highly mature in every area of their operations (most are located within the first two maturity stages: see Harmon 2009). Indeed, well established organizations can have immature processes, whilst new organizations can have highly mature processes.

Immaturity	**Maturity**
Passive	Active
Dependent	Independent
Small number of behaviors	Many behaviours
Erratic and shallow interest	Deep and strong interests
Short term	Mid and long term
Subordinate	Equal or superordinate
Lack of self-awareness	Awareness and self-control

Table. 6. OMT Continuum.

OMT is closely linked to the use of Maturity Matrices, which offer structured representation of an organization's transitions through various stages and states of development (see for example Maier *et al*. 2011). In conjunction, OMT and Maturity Matrices can drive the change required to embed TBE in museums, through a constant cycle of reflection leading to a 'roadmap' for enacting change.

Conclusion

In this paper, we have attempted to shed further light on Falk and Dierking's (2000) question 'Why do people go to museums and what do they learn there?'. In doing so, we have examined the intersection between policy, performativity and evaluation via the voices of museum educators from three different countries. In their accounts, evaluation was increasingly accountability- and impact-focused, rarely if ever team based, and often disciplinary. This international context (albeit three European countries) is important, as it highlights how performative agendas not only shape how NF-L is evaluated, but also the values ascribed to why people go to museums and what they learn when they do.

Despite these agendas, TBE can, to an extent, enable educators to 'fight back'. It can produce rich and meaningful evaluative evidence that can benefit learners, educators and museums alike. TBE groups meanwhile can impact positively beyond the process of evaluating NF-L. It is also worth noting that, although this project was set within the museum education space, TBE principles are applicable to any evaluation context.

Of course, this 'fight back' is not unproblematic. There was little or no resourcing or organizational 'buy-in' for TBE and teamworking was difficult to undertake for core activities, let alone evaluation. There was also scepticism towards evaluation (and therefore TBE) from some educators working in a sector facing reduced funding and yet increasing accountability. Nonetheless, the multi-educator and international discourses that have been examined here

offer a compelling argument as to why TBE has a part to play in evaluating NF-L at the local and macro scales.

At the local scale, TBE has the potential for producing rich evaluation evidence and developing teamworking practices. At the macro national and international scales, it can contribute to resetting NF-L evaluation from an accountability and disciplinary 'machine' to a dialogic, democratic and developmental activity. In both cases, TBE has the capability to develop the 'intrinsically-motivated, joyful, open-ended learning power' (Garcia 2012: 47), so powerfully mediated by museum educators.

Acknowledgements

The authors wish to thank the team of botanic garden and museum professionals who developed LtE and the LtE participants, especially 'Julia' and 'Richard'.

Competing Interest

The authors report there are no competing interests to declare.

Data Availability

Due to the nature of this research, participants of this project did not agree for their data to be shared publicly, so supporting data are not available.

Notes

1 Evaluation can develop organizational processes and practices (Jeffs and Smith 1999).

2 NF-L occurs via complex experiences and interactions (see for example, Sefton-Green 2012; Johnson and Majewska 2022, Martin Johnson and Dominika Majewska, 'Formal, Informal and Non-Formal Learning: Key Differences and Implications for Research', Annual Conference of the British Educational Research Association 12-14 September 2023 [conference presentation]. https://www.cambridgeassessment.org.uk/Images/699142-formal-informal-and-non-formal-learning-key-differences-and-implications-for-research.pdf, accessed 21 May 2023).

3 'Educator' describes professionals working in settings including museums, science centres, theatres, galleries and botanic gardens.

4 'TBE model' refers to the theoretical framework underpinning this evaluation approach. 'TBE groups' meanwhile describes those undertaking evaluations.

5 All the informants had roles as educators in museum education departments and responsibility for evaluating NF-L.

6 There is significant evidence highlighting the benefits of teamworking (for example, Schmutz et al. 2019).

7 Department for Education, 'Opportunity for All: Strong Schools with Great Teachers for Your Child', 2022. https://www.gov.uk/government/publications/opportunity-for-all-strong-schools-with-great-teachers-for-your-child, accessed 21 May 2024. White papers are policy documents produced by the UK Government that set out proposals for future legislation.

8 Department for Education, 'Opportunity for All'.

9 Network of European Museum Organisations (NoEMO), 'Money Matters: The Economic Value of Museums', Report for the 24th Annual Conference 10-12 November 2016. https://www.ecsite.eu/activities-and-services/resources/money-matters-economic-value-museums, accessed 21 May 2024.

[10] See also Johnson and Majewska, 'Formal, Informal and Non-Formal Learning'.

[11] Jason Farago, '10 Ways for Museums to Survive and Thrive in a Post-Covid World', The New York Times 2021. https://www.nytimes.com/2021/05/21/arts/pandemic-museums-ideas.html, accessed 21 May 2024.

[12] Sarah Cohn, 'Team-Based Inquiry. Using Evaluation to Improve Educational Experiences', National Science Teaching Association 2019. https://www.nsta.org/connected-science-learning/connected-science-learning-july-september-2019/team-based-inquiry, accessed 21 May 2024.

[13] See Cohn, 'Team-Based Inquiry'.

[14] Scott Pattison, Sarah Cohn and Liz Kollmann, L. 'Team-Based Inquiry. A Practical Guide for Using Evaluation to Improve Informal Education Experiences', 2013: 5. https://www.aam-us.org/wp-content/uploads/2018/01/making-evaluation-practices-work-in-real-time.pdf, accessed 10 November 2022.

[15] Pattison et al., 'Team-Based Inquiry', 6-7.

[16] Cohn, 'Team-Based Inquiry'.

[17] See Pattison et al., 'Team Based Inquiry'.

[18] Learn to Engage LtE (BGCI/Erasmus+ 2016-1-UK01-KA202-024542) developed a suite of professional development modules for museum educators, with the Evaluation module led by the authors. Botanic Gardens Conservation International (BGCI)/Erasmus +, 'Learn to Engage - a Modular Course for Botanic Gardens', Erasmus + (ref. 2016-1-UK01-KA202-024542) 2019. https://erasmus-plus.ec.europa.eu/projects/search/details/2016-1-UK01-KA202-024542, accessed 21 May 2024.

[19] The project received favourable ethical opinion from the university Ethics Committee and all participants provided ethical consent (see British Educational Research Association 2018).

[20] All five themes identified in Phase I were interlinked, for example, the 'Working as a team' theme strongly linked to the 'Resourcing' theme. Similarly, all five themes were present in the data from educators working in Portugal, Italy and the UK.

[21] A detailed exploration as to how OMT can support museums to embed TBE is the focus of ongoing work. However, it is worth noting that tools such as maturity matrices (see Jokela et al. 2006) can be employed to drive the organizational change.

References

Adams, M. (2012) 'Museum Evaluation: Where Have We Been? What Has Changed? And Where Do We Need to Go Next?', Journal of Museum Education, 37 (2) 25-35.

Ball, S.J. (2003) 'The Teacher's Soul and the Terrors of Performativity', Journal of Education Policy, 18 (2) 215-28.

Barata, R., Castro, P. and Martins-Loução, M.A. (2017) 'How to Promote Conservation Behaviours: The Combined Role of Environmental Education and Commitment', Environmental Education Research, 23 (9) 1322-34.

Braun, V. and Clarke, V. (2021) Thematic Analysis: A Practical Guide, London: Sage.

British Educational Research Association (2018) Ethical Guidelines for Educational Research, (4th edn), London: BERA.

Buckley, J., Archibald, T., Hargraves, M. and Trochim, W.M. (2015) 'Defining and Teaching Evaluative Thinking: Insights From Research on Critical Thinking', American Journal of Evaluation, 36 (3) 375-88.

Bulaitis, Z.H. (2020) 'Impact and the Humanities: The Rise of Accountability in Public Cultural Life', in Zoe Hope Bulaitis, Value and the Humanities, 177-240, Cham: Palgrave Macmillan.

Clapham, A. (2023) 'Docility and Dilemmas: Mapping "Performative Evaluation" and Informal Learning', Journal of Education Change, 24 1077-98.

(2024) 'Undiminishing School Governance: Investigating "Governance Maturity Theory" for School Governing Bodies', Journal of Education Policy. https://doi.org/10.1080/02680939.2024.2337031

Clapham, A., Vickers, R. and Eldridge, J. (2016) 'Legitimation, Performativity and the Tyranny of a "Hijacked" Word', Journal of Education Policy, 31 (6) 757-72.

Cohn, S., (2019). Team-Based Inquiry. Using Evaluation to Improve Educational Experiences. Available at: https://www.nsta.org/connected-science-learning/connected-science-learning-july-september-2019/team-based-inquiry

Crosby, P.B. (1979) Quality is Free, New York: McGraw-Hill.

Curioni, S. and Rizzi, E. (2016) 'Two Realms in Confrontation: Consensus or Discontinuity? An Exploratory Study on the Convergence of Critical Assessment and Market Evaluation in the Contemporary Art System', Museum Management and Curatorship, 31 (5) 440-59.

Dahler-Larsen, P. (2011) The Evaluation Society, Redwood City: Stanford University Press.

Falk, J. and Dierking, L. (2000) Learning from Museums. Visitor Experiences and the Making of Meaning, Plymouth: Rowman and Littlefield.

Garcia, B. (2012) 'What We Do Best', Journal of Museum Education, 37 (2) 47-55.

Harmon, P. (2009) 'Process Maturity Models', Business Process Trends, 2 (5) 1-5.

Jeffs, T. and Smith, M.K. (1999) 'Informal Education', Conversation, Democracy and Learning, Ticknall: Education Now.

Johnson, M. and Majewska, D. (2022) Formal, Non-formal, and Informal Learning: What Are They, and How Can We Research Them?, Cambridge University Press & Assessment Research Report, Cambridge: Cambridge University Press.

Jokela, T., Siponen, M., Hirasawa, N. and Earthy, J. (2006) 'A Survey of Usability Capability Maturity Models: Implications for Practice and Research', Behaviour & Information Technology, 25 (3) 263-82, https://doi.org/10.1080/01449290500168079

Kluge-Pinsker, A. and Stauffer, B. (2021) 'Non-visitors: Who Are They and What Should We Do About Them?, Journal of Museum Education, 46 (1) 61-73.

Kubarek, J. (2015) 'Building Staff Capacity to Evaluate in Museum Education', Journal of Museum Education, 40 (1) 8-12.

Kubarek, J. and Trainer, L. (2015) 'Empowering Museum Educators to Evaluate', Journal of

Museum Education, 40 (1) 3-7.

Kvale, S. and Brinkmann, S. (2009) Interviews: Learning the Craft of Qualitative Research Interviewing, London: Sage.

Lyotard, J.F. (1984) The Postmodern Condition: A Report on Knowledge, Manchester: Manchester University Press.

Maier, A.M., Moultrie, J. and Clarkson, P.J. (2011) 'Assessing Organizational Capabilities: Reviewing and Guiding the Development of Maturity Grids', IEEE Transactions on Engineering Management, 59 (1) 138-59.

Pattison, S., Cohn, S., and Kollmann, L. (2013). Team-Based Inquiry. A Practical Guide for Using Evaluation to Improve Informal Education Experiences. Retrieved 10 November 2022 from: https://www.aam-us.org/wp-content/uploads/2018/01/making-evaluation-practices-work-in-real-time.pdf.

Preskill, H. (2011) 'Museum Evaluation Without Borders: Four Imperatives for Making Museum Evaluation More Relevant, Credible, and Useful', Curator, 54 (1) 93-100.

Schmutz, J., Meier L. and Manser, T. (2019) 'How Effective is Teamwork Really? The Relationship Between Teamwork and Performance in Healthcare Teams: A Systematic Review and Meta-analysis', BMJ Open, 9 (9) 1-16. https://doi.org/10.1136/bmjopen-2018-028280

Sefton-Green, J. (2012). Learning at Not-School A Review of Study, Theory, and Advocacy for Education in Non-Formal Settings, Cambridge: MIT Press.

Wissema, J. (2000) 'Fear of Change? A Myth!', Journal of Change Management, 1 (1) 74-90.

Yin, R.K. (2003) 'Designing Case Studies', Qualitative Research Methods, 5 (14) 359-86.

***Andrew Clapham**, Nottingham Trent University, Nottingham, United Kingdom.

****Raquel Barata,** Head of Education, Centre for Ecology, Evolution and Environmental Changes, National Museum of Natural History and Science, University of Lisbon, Portugal.