Towards an "Evaluation Dilemmas Model" – designing an evaluation scheme for a European capital of culture

Hans-Peter Degn

School of Communication and Culture, Aarhus University, Aarhus, Denmark

Steven Hadley

National University of Ireland Galway, Galway, Ireland, and

Louise Ejgod Hansen

School of Communication and Culture, Aarhus University, Aarhus, Denmark

Abstract

Purpose – During the evaluation of European Capital of Culture (ECoC) Aarhus 2017, the evaluation organisation rethinkIMPACTS 2017 formulated a set of "dilemmas" capturing the main challenges arising during the design of the ECoC evaluation. This functioned as a framework for the evaluation process. This paper aims to present and discuss the relevance of the "Evaluation Dilemmas Model" as subsequently applied to the Galway 2020 ECoC programme evaluation.

Design/methodology/approach – The paper takes an empirical approach including auto-ethnography and interview data to document and map the dilemmas involved in undertaking an evaluation in two different European cities. Evolved via a process of practice-based research, the article addresses the development of and the arguments for the dilemmas model and considers its potential for wider applicability in the evaluation of large-scale cultural projects.

Findings – The authors conclude that the "Evaluation Dilemmas Model" is a valuable heuristic for considering the endogenous and exogenous issues in cultural evaluation.

Practical implications – The model developed is useful for a wide range of cultural evaluation processes including – but not limited to – European Capitals of Culture.

Originality/value – What has not been addressed in the academic literature is the process of evaluating ECoCs; especially how evaluators often take part in an overall process that is not just about the evaluation but also planning and delivering a project that includes stakeholder management and the development of evaluation criteria, design and methods.

Keywords Evaluation, Cultural events, European capitals of culture, Evaluation dilemmas, Evaluation process

Paper type Research paper

Introduction

Engaging in the evaluation of a European Capital of Culture (ECoC) involves being confronted with specific dilemmas regarding the evaluation process. This article presents these dilemmas in the form of an "Evaluation Dilemmas Model" developed as a part of the evaluation of Aarhus 2017 (Denmark) and tested during Galway 2020 (Ireland). Evaluation of ECoCs is made more

complex by both the scale of the project and the wide range of objectives defined by a multiplicity of stakeholders (Garcia and Cox, 2013; Palonen, 2010; Palmer/Rae Associates, 2004). The generic European goals are supplemented with more specific, local objectives related to the context and strategic priorities of the local, regional and national stakeholders that are the main subsidisers of this event. This makes the evaluation process particularly challenging, because it is impossible to cover everything and answer all stakeholders exhaustively [1].

The European Commission has no specified, extensive evaluation scheme to cover the impacts of an ECoC. As a result, evaluation processes are set up and defined locally – resulting in very different formats, scale, organisations, etc. In Aarhus, it was stated in the bid book that the university should be responsible for the evaluation. In Galway, it was a tender very late in the process. As academics involved in the evaluation procedures in the two ECoCs, we explore these challenges and demonstrate how the "Evaluation Dilemmas Model" can function as a framework to address them. Comparison between different capitals of culture is seldomly undertaken yet provides the opportunity to discuss the commonalities and differences between institutional, historical and national contexts.

Our analysis is based on two very different ECoCs. Where Aarhus 2017 was generally seen as a well-managed and successfully delivered ECoC, Galway faced the consequences of the COVID-19 pandemic, was marked by lockdown, management changes and severe budget cuts. Our two cases thus demonstrate that the "Evaluation Dilemmas Model" is relevant for very different ECoCs. Whether the model is relevant for other ECoCs as well remains untested. Despite the fact that several evaluations of ECoCs have been conducted, either by private research companies (Zentrum für Kulturforschung und ICG Culturplan, 2021), the delivery team itself (Turku, 2011 Foundation, 2012), or by the local university (Garcia *et al.*, 2010; Berg and Rommetvedt, 2009; Bergsgard and Vassenden, 2009; Degn *et al.*, 2018a, b), the focus on the evaluation process and its inherent dilemmas remains unexplored.

The purpose of this article then is to present, discuss and reflect upon key dilemmas inherent in the evaluation of European Capitals of Culture. We use different data to document and map the dilemmas involved in undertaking a programme of evaluation in two different European cities, Aarhus and Galway, combining our own project experience from Aarhus 2017 with a series of key informant interviews in Galway. We consider both the development of the "Evaluation Dilemmas Model" and the potential for wider applicability in the evaluation of large-scale cultural projects. Whilst the model is derived from practice, our analysis is theoretically informed by evaluation theory, mainly that of the Danish Professor in Political Science, Dahler-Larsen (2006a, b, 2009, 2012, 2013).

The starting point for the model is the evaluation of Aarhus 2017. Here, the strategic objectives of the ECoC were formulated in the application phase instigated by Aarhus Municipality in 2007. The application thus included a plan for the evaluation to be conducted by Aarhus University in a way that matched the overall approach to, and theme of, Aarhus 2017: rethink. After being designated ECoC 2017 in 2012, the delivery organisation, The Aarhus 2017 Foundation, was established. Together with Aarhus University, Aarhus Municipality and Central Denmark Region, they established the project organisation *rethinkIMPACTS 2017* with responsibility for the evaluation. The partnership-based organisation meant that all the main stakeholders of the evaluation – and thus also the different interests – were represented in the steering committee. These interests needed to be addressed before the evaluation team could develop their evaluation model. It was as a part of this process, that the nine dilemmas were formulated (see Table 1).

These dilemmas stem directly from the process of aligning the somewhat conflicting interests amongst the main stakeholders of Aarhus 2017, supplemented by experiences from Galway (2020). Based on the Aarhus process, the dilemmas have been presented in talks and discussions with other evaluators and academic peers. During these conversations, it became evident that the dilemmas were generally recognisable. Time and time again when setting up evaluations, one has to consider these dilemmas and mitigate potentially conflicting interests

Dilemma		
1	Breadth	Depth
2	Summative	Formative
3	Evaluation of results	Evaluation of processes
4	Existing methods and indicators (replication)	New methods and indicators (development)
5	Fixed evaluation scheme	Dynamic evaluation scheme
6	Research-based	Researcher-based
7	Commissioned	Independent
8	Analysis of outputs	Analysis of outcomes/impacts
9	Short-term	Long-term

Source(s): Developed by Hans-Peter Degn based on the design phase of the Aarhus 2017 evaluation. Subsequently tested on the Galway 2020 evaluation

to ensure matching (or aligned) expectations. The purpose of this article is not to provide a fixed approach for the "perfect" evaluation scheme. These dilemmas do not offer an answer as to where exactly to position, or how to configure, an evaluation. What they provide is a framework to clarify and assess the competing priorities which need to be considered when planning an evaluation [2].

The article will now offer an analysis and discussion of these dilemmas in the context of an "Evaluation Dilemmas Model" and will examine both the way *rethinkIMPACTS 2017* approached the evaluation of Aarhus 2017 and the perceived suitability and applicability of the model for Galway 2020.

1. Breadth vs depth
This dilemma debates whether full coverage is preferable to a selective, but more in-depth approach producing more substantial knowledge on specific aspects of the project. This dilemma is relevant due to the complexity of the objectives of an ECoC. Covering all possible aspects of the project in an evaluation would inevitably result in the loss of richness and depth which might result from a more targeted and narrower focus.

In Aarhus, a first step was to deduce from the original six strategic objectives, specified in the application, a set of evaluation criteria that were (more or less) evaluable. As each objective comprised several aspects to (potentially) evaluate, they had to be broken down into individual elements, each of which contained only one evaluation criterion. This process resulted in a list of 43 evaluation criteria, a number that in itself indicates the proportions of full coverage. For example, strategic objective no. 3 was:

> 2017 should employ creativity, innovation, knowledge and experimentation to fuel human development and economic growth.

In order to break this down, an understanding of the concepts of "creativity", "innovation", "knowledge", "experimentation" and "human development" needed to be developed and agreed upon. *rethinkIMPACTS 2017* chose an inductive approach, calling for workshops where stakeholders and actors involved in the ECoC were invited to interpret the objectives. Thus, the translation into evaluation criteria was based on the understanding amongst key actors.

Alternatively, given that all these concepts have been discussed extensively in academic literature, a research-based definition of such evaluation criteria would be a time-consuming task involving both a discussion with stakeholders about the understanding of the concepts and a theoretically informed clarification. On top of that, selection of indicators of, e.g.

"economic growth" would need to be made: Would increase in turn-over in the tourist industry be a reasonable measurement given one could expect a causal relation between that and the fact that the city was European Capital of Culture? Would an increase in the number of companies in the creative industries be indicative, and when should that be measured? Or would unemployment rate or macroeconomic growth be the appropriate measure, given that might have been the expected outcome when local politicians approved the plan to become ECoC? All these questions demonstrate that an in-depth, full coverage evaluation may not be an option, and that each ECoC needs to weigh the two possibilities.

In Aarhus, the stakeholders represented in the steering committee chose for "scattered in-depth". This allowed for prioritisation amongst the 43 evaluation criteria, and to go in-depth with those selected. However, at the same time, the evaluation had to ensure that the selected criteria covered aspects from all six strategic objectives. In Galway, the same priority was made: "it is very important to have clear information on how funding was spent" (G6) [3] and to be able to document whether the ECoC met its objectives. From a funders' perspective, breadth is often prioritised, whereas other stakeholders might have interests in depth. This is most easily defined as the difference between audit and evaluation. Audit requires that public funds are spent according to the legislative and administrative requirements set out in delivery and funding agreements. Stakeholder engagement is not, however, limited to audit concerns and there is an overall sense of both audit and evaluation being important, such as "within audience development, the capacity building of the sector" (G6). Breadth is required to produce "a holistic approach" (G2) which affords a range of indicators designed to appease a range of stakeholders, but such an approach engenders a focus on "superficial top line numbers" (G2) and is regarded as a concomitant part of "this obsession with trying to prove success by numbers" (G2). Dilemma no. 1 demonstrates the need to find a balance between the wishes of stakeholders in any specific case, since "breadth" and "depth" both have advantages and disadvantages.

2. Summative vs formative

Dahler-Larsen (2006b) operates with seven purposes of evaluation, stating that the two most common are "control" and "learning". The balance between the two is closely linked to the prioritisation between a summative and a formative evaluation. A summative evaluation sheds light on the project in hindsight; often addressing questions like "how did it go?", "was it worth the effort?". Whereas a formative evaluation takes place during the project and is thus not only retrospective but also addressing questions like "what needs to be adjusted?", "how can we do better?".

A formative approach will thus serve an explicit purpose of learning along the way. Because of the one-off nature of being European Capital of Culture, for an ECoC to obtain learnings from its own evaluation, the evaluation process must include a formative approach. Contrary to this, for cultural institutions or repeated projects, a summative evaluation can provide learnings useful for their next iteration. Thus, the summative evaluation might both serve the purpose of control – documenting the results of the project – and provide enlightenment for different actors and stakeholders.

In Aarhus, part of "rethinking" evaluation was an emphasis on formative evaluation. However, both in Aarhus and later on in Galway, formative evaluation became almost immediately problematic – not in theory, but in practice – as soon as the ECoC "mindset" (G1) became focussed on production and delivery. Tied to this pressure, with non-negotiable lead times, is the sense that formative evaluation can raise questions such as "Are you going to hold on to your funding if you decide to do something differently?" (G1) and, from a different perspective, the idea of changes wrought by formative evaluation being seen as "artistic interference" (G1). In Galway, a key knowledge gap was around the question of integrating

formative evaluation into delivery: "what does it say to the marketing team … what does it tell the programme team? But then you're so deep into delivery, it's very difficult to react" (G3). Equally, there was a concern that the summative evaluation may not be strategically aligned and integrated with wider arts strategies across Galway City and County due to gaps in the strategic integration of the ECoC objectives amongst the various stakeholder planning processes.

Summative evaluations and their attendant advocacy telling "the story of the success" were felt by some of the interviewees in Galway to have contributed to an inter-ECoC metric inflation, which both generated a "space race" in terms of expectations and adversely affected smaller cities due to economies of scale. The question of summative vs formative also raised the issue of "inter-ECoC learning" (G2) because "a lot of the same mistakes are being made" (G2). This demonstrate that from the perspective of the European Commission and the ECoC as an EU policy initiative, there is a potential for learning not linked to the in-process formative approach, but to a summative approach including process and learnings. The dilemma of summative vs formative is thus not a dichotomous choice between learning and control, but a choice regarding when and for whom learning is both appropriate and important.

3. Evaluation of results vs evaluation of processes

Linked to the dilemma of formative vs summative evaluation is the question of whether the evaluation should focus on results or processes. A formative approach tends to focus on processes. However, it is not sufficient to only focus on the processes themselves. Furthermore, results gained from projects and preceding phases can provide important contributions to formative processes for subsequent projects and phases.

A dimension of this dilemma is the balance between producers and consumers in the evaluation. A process-oriented evaluation tends to put emphasis on producers, whereas a result-oriented evaluation would tend to combine the two, probably with an emphasis on consumers. The broad range of data collected in Aarhus included both producers and consumers which rendered possible the inclusion of both process and results [4]. In Galway, there was felt to be an overt focus on production as opposed to consumption, which caused problems for the evaluation process. This was felt to be a local instance of a wider, national approach to cultural value and evaluation with a focus heavily geared towards artists rather than audiences. It was also noted anecdotally that several cultural partners queried the necessity of data for evaluation as "the Arts Council has never asked us for this information" (G2).

> Ideally, the process is important because that's where knowledge transfer happens and that's where, you know, you also create a legacy, I suppose. So you understand you've built a capacity in the city and that's really the aim of a European Capital of Culture. And it's not just, you know, a festival that descends on the city, but that it's strategically working. (G3)

A key distinction in this dilemma is that within community programme processes, knowledge can be both transferred and retained locally. With large-scale events, where many of those involved are "bought in", this is self-evidently less likely. This raised the interesting question as to whether the ECoC was a unique event, or whether it was simply a unique cluster of non-unique events. In Galway, it was felt that if process evaluation meant "how things happen in the year and the sorts of ways that what seems to work or what doesn't work, of course, that's absolutely fundamental." (G2). This approach was felt to be especially important in the context of rural engagement programmes (see, e.g. G2020 Small Towns Big Ideas [5]) where there was significant applicability across future ECoCs with large rural populations. Again,

however, the issue of knowledge transfer came up, with the feeling that "a really ridiculous, continual reinvention of the wheel" (G2) was a more likely scenario.

From a delivery perspective, it was clear that "the how becomes more important than the what" (G5) and that process evaluation was not solely an operational concern. But from a policy and stakeholder perspective there is inevitably a sense that results – in terms of visibility and impact – "are your key measures" (G6) and that process "doesn't get the level of attention that maybe it should" (G6) such that although Galway 2020 "has been monitored, which would speak to the processes side . . . the evaluation of results is key actually". Equally important in this dilemma, and discussed later in the article, was the "opportunity to explore, change and fail" (G1) (see also Jancovich and Stevenson, 2020).

4. Existing methods and indicators (replication) vs new methods and indicators (development)

In both an academic and policy context, there is an increasing demand for new approaches to evaluation and new indicators of the value of culture (Myndigheten för kulturanalys, 2012; Reeves, 2002). For large parts of the wide-ranging impacts that ECoCs are trying to cover, there are few or no good methods or indicators available. The development of new indicators and methods can therefore appear a priority for ECoCs. One example of why new indicators and methods are needed was strategic objective no. 5 of Aarhus 2017:

> 2017 should support the development of open and vibrant urban environment which creates a common and cohesive framework whilst allowing and encouraging local, social and cultural diversity (Aarhus Candidate European Capital of Culture 2017, 2012).

This aim of using culture as a tool for the regeneration of cities is quite common for ECoCs but the way in which it has been evaluated is inadequate (Garcia and Cox, 2013, p. 132). Simple measures like physical changes of urban spaces or the flow of people in specific places in the city might be applied as basic indicators, but that tells us little to nothing about "vibrancy" or "diversity". While developing new approaches to evaluation was a part of *rethinkIMPACTS 2017*, the main part of the evaluation was based on well-known data sources mixing qualitative and quantitative methods. It was felt that a focus on the development of new methods and indicators was a gamble because it might fail; when developing new methods or indicators there is an immanent risk that these might turn out to be, for instance, too difficult to implement or less appropriate than already existing and tested methods.

Given the high-profile characteristics and one-off nature of an ECoC, there is limited scope for such dead-ends. This applies equally to the evaluation, and there is scarce time for redevelopment of unsuitable methods or indicators. For these reasons, only a minor part of the evaluation of Aarhus 2017 took up the challenge of developing new methods and indicators – covering, e.g. urban environment (Jensen *et al.*, 2015) and a new cultural segmentation model for a broad range of cultural activities (Degn and Hansen, 2022).

In the case of Galway 2020, given the context of COVID-19, new methods were inevitably required to evaluate online activity and engagement. It was felt, however, that these methods tended to be "technologically pre-determined" (G2) and that irrespective of the validity, scope or scale of the online metrics reported,

> I've never known anyone properly challenge any of these numbers. People say we had two hundred and fifty thousand engagements online or something or users or whatever. They put it in all kinds of Arts Council reports or whatever, and it is taken at face value. (G2).

The question of how to undertake online evaluation – and the concomitant need to upskill cultural organisations to self-administer such processes – is "a massive hole" (G2) which

needs to be addressed by the European Commission and cultural mega-events in general: "How can you do proper evaluation of online digital activity or what it means?" (G2).

In terms of the challenge of developing new methods and indicators, the risk associated with this approach was felt very keenly in Galway, as was the case in Aarhus: "Failure is not possible . . . you are so under scrutiny and being attacked" (G3), and:

> "One size fits all" never works, and it isn't very interesting . . . I think you need to approach it on its own terms, and you do need to be able to cut data and experience across a lot of different levels . . . you need to be using a multitude of methodologies and approaches to get that really textured experience and it's hard to do because the financial resources are finite. (G4)

This raises the question as to the discrepancy between what the producers want to know and what the funders want to know. For artists and producers, there is predominantly a sense that "Quality of experience is more important than the number of people" (G5). However, quality of experience is not as easily measured as numbers of people, which demonstrates that the balance between existing and new indicators is not just a question of whether to engage in risky methodological experiments, but also a question of the objectives and impacts that should be included in the evaluation.

Given the frequently performative nature of claims for excellence and impact in the arts (Belfiore, 2016), the lack of adequate financing for evaluation as well as legacy planning and the political desire for success can lead to the fallacy of *post hoc ergo propter hoc*. Post hoc is a particularly tempting error in cultural evaluation because correlation appears to suggest causality. The fallacy lies in drawing a conclusion based exclusively on the order of events, rather than taking into account other factors potentially responsible for the result that might rule out the connection. The socially embedded nature of cultural engagement makes longitudinal analysis as prone to this bias as political need makes it attractive.

In the context of evaluation, it is important to consider both what is learnt locally (in the place-based context of delivery) but also what is learnt about ECoCs *per se*. Again and again, the issue of failure (or of being allowed to fail) was prominent. The point was made that, whilst research and development is an accepted component of artistic development and creation, evaluation always needed to be done "right first time" (G1). This expectation pulls towards the "replication" side of this dilemma, whereas the value of the "development" side is that the result might enable new ways to capture more complex outcomes and impacts that cannot be measured with existing methods.

5. Fixed evaluation scheme vs dynamic evaluation scheme

Linked to the dilemma of new and existing methods is the dilemma of a fixed vs a dynamic evaluation scheme. With a project of the dimension of an ECoC, it is expected that the project will develop during the lead-in time as well as throughout the year itself. An important part of the evaluation is an ability to reflect on and potentially incorporate these changes in the evaluation process. However, as the European Commission (2018, p. 12) states, "Timely implementation of the evaluation helps to ensure that appropriate organisational arrangements are put in place, that the funding is planned and time is allocated to establish data collection and analysis frameworks, as well as the baseline position".

The main advantage of a fixed evaluation scheme is the potential to compare across time. But the absence of national baseline data can be a considerable hindrance for the timely development of a fixed evaluation scheme. This was the case in Aarhus as well as in Galway, where "the absence of baselines was an ongoing source of concern and debate" (G4) and "there was none nationally. And that was a real difficulty, actually, because where do you start? How do you benchmark? How do you really assess success and failure if you don't have a starting point?" (G5). In Aarhus, the evaluators were partly able to compensate for the lack

of existing data by starting to gather data in the years preceding the event, whereas the late start of the evaluation in Galway made this particularly problematic. There was a clear recognition, from designation in 2016, that "the collection or identification of relevant baseline data" (G6) was both important and absent.

However, baseline-based, fixed evaluation schemes might also raise expectations regarding the development potential of the ECoC. This was the case in Aarhus 2017, where the ECoC was expected to raise cultural participation in the regional population. The – reasonable – explanation why this was a too ambitious objective, was given far less attention than the fact that Aarhus 2017 "failed" in this (see Hansen and Degn, 2022). In Galway, the obvious reason to have a more dynamic approach to evaluation was the COVID-19 pandemic that totally changed the possibility and conditions for the ECoC. Any analysis of baseline data had to take this game changing element into consideration.

6. Researcher-based vs research-based

It seems immanent that evaluations should be evidence-based. This relates to how the evaluation (methods, indicators, theoretical explanations, etc.) is backed by rigorous research. Accordingly, this dilemma considers whether evaluation should be research-based as opposed to merely researcher-based. By this we mean to consider the role of the (usually local) university in the evaluation process. Engagement with a university implies that academic researchers will conduct the evaluation. Compared to either a self-evaluation or an evaluation conducted by a private consultancy firm, the university-conducted evaluation externally signals high quality knowledge and critical, independent thinking. However, the mere fact that the evaluation is conducted by a university does not mean that such attributes are delivered.

One challenge in relation to the delivery of a research-based evaluation is the broadness within the strategic objectives of an ECoC and consequential inclusion of different impact areas, which means that a range of researchers from different disciplines would need to be engaged. *rethinkIMPACTS 2017* stimulated this engagement in several ways. For example, seed money was provided to researchers to engage in either pilot studies or the development of methods that could be used in the evaluation. In addition, researchers from a broad range of academic disciplines were invited to take part in the development of the evaluation. Workshops were delivered to specify and operationalise the overarching impact areas and the strategic objectives related to each of them. Later in the process, engagement of researchers was linked to specific subparts of the evaluation. This is one attempt to fulfil the ambition to conduct a research-based evaluation, avoiding reducing the value of an evaluation conducted by a university to the value of a researcher-based evaluation. In that scenario, the legitimacy of the evaluation may be linked mainly to the academic title of the evaluators rather than their actual insights within the specific areas examined.

Inevitably, there is a discrepancy between an external, idealised view of what a university brings to the evaluation process, and the reality. Equally, there is a significant practical question as to how to ensure complementarity (in terms of outputs, timelines, access to data) between academic research and management consultancy. As experienced in Aarhus, there is frequently an external assumption that a university will have all and any pre-requisite skills "in-house" to undertake evaluation of major cultural events. Such a viewpoint fails to understand the internal variation amongst subject strengths within universities – for example, the existence or lack of academic research skills in cultural policy, arts management and a broader understanding of the creative and cultural industries field beyond traditional university subject areas such as drama, film, etc. In Galway, the university was thus understood to bring "rigour and objectivity" (G6), even though it is clear that "not every university would be equally qualified" (G6) to undertake the complexity of an ECoC evaluation project. Where the university can add value is by providing "a helicopter view and

outside opinion" (G5). A problem with the university as evaluation partner approach comes when the university "has no idea of evaluation at all" (G1).

Key to this dilemma is the temporality and materiality of the university as an institution – it is both of "place", being a symbolic and important institution in the host city and exists/ remains both pre and post-ECoC. This provides a much-needed degree of consistency and constancy to the evaluation process. Questions were raised as to whether the partner university needed to be "local" (for example the Rijeka 2020 evaluation was conducted in partnership with University of Sarajevo) and how both the physical and metaphorical "distance" of the university might affect research objectivity and the inevitable stake a locally situated university might (problematically) have in the evaluation process. This is especially the case when the local university is also involved in the cultural programme. Even though this was the case in Aarhus, it was to a much lower degree than, e.g. in the UK Cities of Culture, Hull 2017; Coventry 2021. Whilst academic researchers might nobly consider themselves to be objective, the university as a situated institution is inevitably political.

Whilst the university's engagement may be an "automatic assumption" (G4) and "part of the accepted process" (G4) it is important to acknowledge that the prestige of association is bi-directional: the university gains prestige by being associated with/assisting in the delivery of the ECoC alongside the status and credibility perceived to accrue to the evaluation process via the university's involvement. Whilst there was general recognition that the university "adds credibility" (G2) there were practical concerns to be taken into consideration – key was the difference between what one Galway respondent articulated as "practical research" (G2) – research with a quick turnaround and immediate impact – as against the longer-term focus, timelines and delivery of academic research. This was also articulated as a difference between the culture company (ECoC delivery organisation) being focussed on delivery, whilst academics treat the ECoC as an object of study.

In terms of resource (in this case that of knowledge and labour) there is a sense in which the university is seen to provide "free labour" to the ECoC. In Aarhus, the university contributed with half of the costs of running *rethinkIMPACTS 2017*, giving it the status of a strategic project for the university. In Galway, this was, "in part why the process in Galway was so painful" (G4), due to the very time and labour-intensive process involved in extracting this "free labour" from the university in the first place.

7. Commissioned vs independent
Every evaluation tends – to some degree – to be commissioned work, but the way in which it is commissioned is key in this dilemma. The primary issue is independence from the contracting authority funding the evaluation, from the part being evaluated as well as other stakeholders. In the case of an internal evaluation, this is particularly obvious; there is no doubt that the evaluation represents a specific and intentionally controlled perspective on the ECoC. But even in the case of an external evaluation, there is a question of independence. There are several dimensions to the question, for instance the question of publishing and communicating evaluation results. Questions of who is defining the evaluation criteria, and whether the evaluators can independently decide how to interpret, operationalise and prioritise the evaluation criteria are pivotal for the design of the evaluation and thus the selection of sources that the evaluation will build upon. Another aspect is organisational and regards the power which the ECoC organisation or key stakeholders can potentially exercise over the evaluators, especially if they feel threatened by the results.

In the case of Aarhus 2017, the three main stakeholders were represented in the steering group of *rethinkIMPACTS 2017,* thereby influencing the ongoing evaluation process, but with a stated respect of the independence of the evaluators. As the evaluators were located at the university, there was also a certain level of peer pressure to present non-biased data and

an evaluation inclusive of critical perspectives. This was not necessarily in the interest of the delivery organisation and post-event two different reports were published: In April 2018 the delivery organisation published their own report on the short-term impact of Aarhus 2017, based on their own monitoring data (Simonsen, 2018) telling the success story, emphasising the positive results. In December 2018, *rethinkIMPACTS 2017* presented their more nuanced evaluation combining short- and long-term effects and unfulfilled potentials (Degn *et al.*, 2018a, b).

The Galway 2020 evaluation was an external evaluation, a decision taken to ensure independence; "If you have integrity, you don't commission the results . . . evaluation at the end of the day, its purpose is to learn. So if you don't want to learn, then you're commissioning results." (G6) Inevitably, there is always a danger that evaluation will be compromised to some degree. There is still debate operating on some idealised level of thinking, however, "I think if you are clear where the compromise is, how you were commissioned, what your job was, what your budget was, who you spoke to, the question, you know, you have to put it all out there and say, this is what we did" (G1).

There was belief that the Galway 2020 evaluation was "very fair, very open and actually very transparent and very partner-led" (G5). Nonetheless, the external pressure to ensure successful evaluation results cannot be over-stated. For example, for Galway 2020 there was the sense that "the fierce local media would just kick it in the balls the whole time" (G4). This perceived inability to be allowed to fail (and therefore to consider experimentation) resulted in the sense that, "I mean, failure is generally not accepted. And it would be really interesting, I mean, it would be more interesting, wouldn't it, really to be able to be open minded enough and open to learning enough to be able to recognise failure (G4). What is expressed here is a general dilemma of evaluation in the cultural sector (Jancovich and Stevenson, 2021), but one that might be particularly relevant to ECoCs since they can be considered projects that are "too large to fail". Taking this into account, the value of an independent evaluation – documenting both the successes and the failures of the ECoC – potentially provides added value to external stakeholders because of the credibility of what is seen as a non-biased conclusion.

8. Analysis of outputs vs analysis of outcomes/impacts

This dilemma asks at what level the evaluation should engage. Implicit within this issue are further questions as to how to address the issue of causality regarding outcomes/impacts (especially longitudinally). Strategic objectives (included in, but not limited to, ECoC bid books) frequently offer grandiose visions of change at local, regional and national levels but the question of what the delivery company can (realistically) claim responsibility for and/or be accountable for is an important one. In many senses, this dilemma asks whether the evaluation should focus on basic and immediate outcomes (e.g. number of audience visits/ how the ECoC is portrayed in the media) or on the impacts (e.g. a more general change in audience behaviour or cultural interest/how citizens' attitudes towards the ECoC project develop over the years).

In Aarhus, responsibility for data gathering was divided between the delivery organisation, responsible for the continuous gathering of monitoring data, and *rethinkIMPACTS 2017*, being responsible for gathering evaluation data. This pushed the evaluation towards a focus on both outcomes and impacts. Responses to this dilemma in Galway articulated the need to move beyond an econometric framework and towards "meaningful" ideas of "inclusive growth, you know, every nation in the UK is articulating it and approaching it in a slightly different way" (G1), such that: "My favourite metric at the moment is, 'are there now people that you can rely on in your community?'" (G1). This was also partly a response to the sense that the Bid Book "is a sales pitch" (G3) and there is a

concomitant need to address the discrepancy between ambition and reality with a meaningful set of output/outcome metrics.

9. Short term vs long term

Linked to dilemma 8, is the dilemma of short over long-term effects and how the evaluation balances these. Whereas ECoCs often include long term objectives, measurement of such objectives is challenged by the short-term timeline for reporting. The interest in long-term effects is shared by the European Commission that states that "there is still a shortage of a coherent evidence-base to better grasp the benefits of being an ECoC, especially its medium-to-long term cultural, social and economic legacy in host cities" (European Commission, 2018, p. 5). Whilst the European Commission's new evaluation model from 2020 seeks to embed evaluation more firmly within the ECoC, it does, however, generate an operational issue in terms of the lifespan of the delivery company as against the lifespan of the evaluation. For Aarhus, the project period of *rethinkIMPACTS 2017* ended at the end of 2018, and in Galway, the reporting requirement was timelined for November 2021. Both our cases demonstrate that there is a need to transfer out the evaluation contract beyond the life of the delivery company. Such a necessity "just shows you how defunct the system is, unfortunately" (G3).

This dilemma reflects back the extent to which stakeholders – in the cases of Aarhus 2017 and Galway 2020 the city and the region/county – regard the ECoC as a strategic priority and thereby seek to maximise its potential within broader, longer-term contexts. Regrettably, the sense in Galway seems to be "that ownership is not there. You know, there's no ownership of this really being a strategic project for the city" (G3). In Aarhus, five years on since the ECoC year, there is currently still no plans to do follow-up studies of the long-term impacts.

This dilemma also raises important questions of the legacy of ECoCs (see, e.g. Ganga, 2022). This remains a vexed question for many ECoCs (and equivalent policy initiatives such as the UK City of Culture). Beginning the evaluation with a short-term perspective might hinder follow-up evaluations focussing on the long-term perspective. Structural deficiencies and weaknesses in the initial evaluation have inevitable repercussions for discussion and enactment of the evaluation of legacy.

Whilst the reality of legacy delivery in Galway remains an open question, there are stakeholders who still question the return on investment, with politicians seeking an independent review of spending on the Galway 2020 European Capital of Culture programme (The Times, 2021). The need to focus on delivery, i.e. on operational necessity, was such that a budget of €1m ring-fenced for the legacy of Galway (2020) European Capital of Culture was diverted to pay for operational and programme costs. The news came (Connacht Tribune, 2020) as Galway City Council confirmed it was withdrawing a €700,000 funding commitment pledged for the project as a result of the severe impact of COVID-19 on its revenue streams.

Discussion

Evaluation of ECoC is challenging for many reasons. Key to the overarching process are the many stakeholders and the ambitious objectives aiming to create a broad range of impacts, but also the time span from the application phase, through the delivery phase to the legacy phase. The different nature of these phases influences the evaluation process. During the application phase, the city is competing for the title, forcing the plans and objectives to be competitive and thus (over)ambitious. In the latter phase, the delivery organisation needs to operationalise the way in which the project and the objectives were described in the original application – a process that also means reiterating the project from something that could win the title of European Capital of Culture into something that is deliverable and that the delivery organisation will be held accountable for (For an analysis of Aarhus 2017s transition from

applicant to designated ECoC see Hansen and Laursen, 2015) for a broad discussion of Galway's development see Collins, 2020; Galway 2020 was, due to COVID-19, more than perhaps any prior ECoC forced to change and develop the project:

> The process of it is all about reaching for the stars and building up and creating huge expectations [. . .]. But then, you know, the reality is that it often happens that some are not delivered and it causes huge falls. So I think between the rhetoric, which is what the modern world aspire to, and the reality, and there's probably a bit of work that needs to be done. (G6)

From the interviewees taking part in discussion around Galway 2020, there was a clearly identified need to de-escalate the inflation of claims for ECoCs, "especially in relation to Bid Book consultancy" (G2). Key to this is a suggested structural flaw in the overall ECoC process, given that the bid book stage – as a competitive process – inevitably encourages a level of rhetoric (what Boland *et al.*, 2018, p. 15 refer to as a "propensity to exaggerate") that can be very difficult (if not impossible) to live up to. This has frequently, as in the case of Galway, resulted in a questioning of the legal and contractual status of the bid book, as host cities struggle to reconcile available resources with what was promised in terms of delivery.

From an evaluation perspective, the transition from the application phase to the delivery phase is essential. If the strategic objectives in the bid book are inflated and the delivery organisation along the way develops more realistic objectives, then the question arises as to which set of objectives the evaluation should be based on. The Aarhus 2017 Foundation reformulated the original strategic objectives into long-term impacts and developed key performance indicators (Aarhus 2017, 2015, pp. 14–15). These covered far less than the original objectives from the bid book, which the delivery organisation was still held responsible for by the European Commission. In the first report from the Monitoring and Advisory Board (2014, p. 5), it was emphasised that Aarhus 2017 was obliged to base their project on the bid book, and the Danish Ministry of Culture also asked the Aarhus 2017 Foundation to report back on the results of the project based on the original strategic objectives. The evaluation criteria of *rethinkIMPACTS 2017* were based on the original six strategic objectives but were developed in close collaboration with a variety of stakeholders. The "Evaluation Dilemmas Model" helped negotiate the evaluation design as a part of this transition process.

Conclusion

The "Evaluation Dilemmas Model" identifies nine dilemmas based on experience from Aarhus 2017 and subsequently tested on Galway 2020. The model is relevant and useful for the evaluation of large-scale cultural events like ECoCs, and also has much wider applicability. The process of developing an evaluation of an ECoC is by no means a well-structured or clear process, and the model is a useful tool for addressing many of the challenges. Implicit in the model is the understanding that each dilemma involves a balancing act rather than a clear-cut either/or dichotomy. Our intention with this article has been to describe, share and reflect upon this ambiguity, believing that this is the only proper way to address it. A part of that is the conviction that our processes of engaging in close cooperation with external partners are not unique. The expectations of an evaluation process are typically linked to the agendas of evidence-based policy and thus on the desire for a firm knowledge base for future decisions (Krogstrup, 2011; Van den Hoogen, 2012). For this reason, evaluations are often expected to provide clear answers and recommendations based on a systematic and clearly outlined evaluation process. Using a metaphor of Dahler-Larsen (2006b, pp. 11–14), the evaluation process can either be a ride on a *model railway* (neatly organised, route planned and predicted, end station known) or a *bumper car ride* (in which the trajectory is unpredictable, each participant needs to find their own route and expect to be run into by some of the other participants during the ride). As the European Commission report (2018, p. 13) ruefully notes,

The balance between pressures to demonstrate quick results and the need to undertake thorough analysis and quality evaluation should be weighed carefully and taken into account at the planning stage.

Recent literature reflecting on the valuing and evaluation of culture with newer methods such as multi-criteria analysis (Getz, 2018; McGillivray and McPherson, 2014) illustrates the complexity of the evaluation of ECoCs. Following this, we argue that the use of the Evaluation Dilemmas Model as both a heuristic and reflective tool can add value to the process. It is our belief that use of the "Evaluation Dilemmas Model" during the evaluation and in the phase of developing the entire project can be of great benefit. As stated at the beginning of the article, the model does not give a clear answer about what the "right" balance is: that answer is specific to the individual cultural event in question. This also means that any framework for evaluation is only ever tentative. Alongside residual resource pressures and inevitable shifts (whether artistic or operational or both), evaluation must always be *responsive* to the changing organisational and operational context within which it operates. We conclude that the "Evaluation Dilemmas Model" is highly valuable in cultural evaluation at the level of the ECoC. We would also argue that it has significant potential for wider applicability in the evaluation of cultural projects way beyond ECoCs. The resources for evaluation are limited and how they should be spent is an ever-present concern. In reality, resource – of which there is never enough, irrespective of type – is a cross-cutting theme of the "Evaluation Dilemmas Model". Nonetheless, by debating and addressing the dilemmas presented in this article early on in the evaluation process, there is a good chance that resources will be spent more wisely.

Notes

1. Several evaluation models such as the ones presented by Parkinson and Bianchini (1994), ECoC Policy Group (2010), Getz (2018) and McGillivray and McPherson (2014) address the complexity of impacts of ECoCs and some of the inherent dilemmas related to ECoC itself. However, this is not the focus of this article. The "Evaluation Dilemmas Model" addresses dilemmas that occur as a part of the evaluation process.

2. The order of the dilemmas does not reflect any kind of chronology or prioritisation. However, it does reflect that some dilemmas do have some affinity to each other.

3. Altogether six interviews were conducted in Galway representing the delivery organisation, the evaluators and the authorities. These interviews are numbered G1-G6 during the article.

4. Especially the thematic report on the organisation of Aarhus 2017 focussed on process (Degn *et al.*, 2018a, b).

5. https://galway2020.ie/en/strand/small-towns-big-ideas/

References

Aarhus 2017 (2015), "European capital of culture Aarhus 2017", Strategic Business Plan, Aarhus, available at: http://www.aarhus2017.dk/sites/default/files/pdf/strategic_business_plan.pdf

Aarhus Candidate European Capital of Culture 2017 (2012), "Application, Kulturforvaltningen Aarhus Kommune", Aarhus.

Belfiore, E. (2016), "Cultural policy research in the real world: curating 'impact', facilitating 'enlightenment'", *Cultural Trends*, Vol. 25 No. 3, pp. 205-216.

Berg, C. and Rommetvedt, H. (2009), *Stavangerregionen Som Europeisk Kulturhovedstad – Slik Innbyggerne Ser Det*, IRIS, Stavanger.

Bergsgard, N.A. and Vassenden, N. (2009), *Stavanger-regionen Som Europeisk Kulturhovedstad – Slik Kulturaktørene Ser Det*, Stavanger: IRIS.

Boland, P., Murtagh, B. and Shirlow, P. (2018), "Neoliberal place competition and *culturephilia*: explored through the lens of Derry~Londonderry", *Social & Cultural Geography*, Vol. 21 No. 2, pp. 1-22, doi: 10.1080/14649365.2018.1514649.

Collins, P. (2020), "And the winner is ... Galway: a cultural anatomy of a winning, designate", *International Journal of Cultural Policy*, Vol. 26 No. 5, pp. 633-648, doi: 10.1080/10286632.2019.1690474.

Connacht Tribune (2020), "Cash black hole grows deeper for Galway 2020", available at: https://connachttribune.ie/cash-black-hole-grows-deeper-for-galway-2020/

Dahler-Larsen, P. (2006a), "Evalueringskultur. Et begreb bliver til. Odense: Syddansk Universitetsforlag".

Dahler-Larsen, P. (2006b), *Evaluering og Magt*, Magtudredningen, Århus.

Dahler-Larsen, P. (2009), *Den rituelle refleksion – om evaluering i organisationer*, Odense: Syddansk Universitetsforlag.

Dahler-Larsen, P. (2012), "Constitutive effects as a social accomplishment: a qualitative study of the political in testing", *Education Inquiry*, Vol. 3 No. 2, pp. 171-186.

Dahler-Larsen, P. (2013), "Evaluering af projekter – og andre ting, som ikke er ting", *Odense: Syddansk Universitetsforlag*.

Degn, H.P. and Hansen, L.E. (2022), "Developing a motivation typology for cultural attendance", Paper for International Conference on Cultural Policy Research, Antwerp, sept. 2022.

Degn, H.-P., Hansen, L.E., Nordentoft, L. and Bak, L. (2018a), "Organiseringen af kulturhovedstadsprojektet Aarhus 2017", *Aarhus: rethinkIMPACTS* 2017, available at: https://projects.au.dk/fileadmin/projects/IMPACT_2017/temarapport_7_rethinkIMPACTS.pdf

Degn, H.-P., Hansen, L.E., Thomsen, N., Nordentoft, L., Xie, J.R. and Hansen, M.D. (2018b), "Aarhus 2017. Before –during – after. A research-based evaluation of the European Capital of Culture project", *Aarhus: rethinkIMPACTS* 2017, available at: https://projects.au.dk/fileadmin/projects/IMPACT_2017/Aarhus2017_before_during_after.pdf

ECoC Policy Group (2010), *An International Framework of Good Practice in Research and Delivery of the European Capital of Culture Programme*, available at: https://ecocpolicygroup.files.wordpress.com/2010/07/ecoc-policy-group_research-framework1.pdf

European Commission (2018), "European Capitals of Culture (ECoC) 2020-2033 Guidelines for the cities' own evaluations of the results of their ECOC", available at: https://ec.europa.eu/culture/sites/default/files/2021-04/ecoc-guidelines-for-cities-own-evaluations-2020-2033.pdf

Ganga, R.N. (2022), "Evaluating cultural legacy: from policy to engaged research", in Wise, N. and Maguire, K. (Eds), *A Research Agenda for Event Impacts*, Elgar, Cheltenham, pp. 229-247.

Garcia, B. and Cox, T. (2013), "European capitals of culture: success strategies and long-term effects", Bruxelles: Directorate-general for internal policies, European Parliament.

Garcia, B., Melville, R. and Cox, T. (2010), "Creating an impact. Liverpool's experience as European capital of culture", Liverpool: IMPACTS08, available at: https://www.liv.ac.uk/media/livacuk/impacts08/pdf/pdf/Creating_an_Impact_-_web.pdf

Getz, D. (2018), *Event Impact Assessment*, Goodfellow, Oxford, doi: 10.23912/978-1-911635-03-1-3897.

Hansen, L.E. and Degn, H. (2022), "No impact on cultural participation? - an analysis of the objective of increasing and widening cultural participation in European Capital of Culture Aarhus 2017", in Cartiere, C. and Schaig, A. (Eds), *The Failures of Public Art and Participation*, Routledge, Oxfordshire.

Hansen, L.E. and Laursen, M. (2015), "From applicant to designated European Capital of Culture", *International Journal of Managing Projects in Business*, Vol. 8 No. 4, pp. 715-731.

Jancovich, L. and Stevenson, D. (2020), "Cultural participation: stories of success, histories of failure", *Conjunctions, Transdisciplinary Journal of Cultural Participation*, Vol. 7 No. 2, pp. 1-8, available at: https://orcid.org/0000-0003-0381-1557

Jancovich, L. and Stevenson, D. (2021), "Failure seems to be the hardest word to say", *The International Journal of Cultural Policy*, Vol. 27 No. 7, pp. 967-981.

Jensen, C.H., Boesen, P., Hansen, L.E., Madsen, S.J., Moth, A., Nybo, N. and Vestergaard, L. (2015), "Kunst og kultur i byrummet under Aarhus 2017. Evaluering gennem et casebaseret metodeudviklingsprojekt", rethinkIMPACTS 2017 reports no. 15, available at: https://projects. au.dk/fileadmin/user_upload/RethinkImpacts_Kunst_og_kultur_i_byrummet_print__.pdf

Krogstrup, H.K. (2011), *Kampen Om Evidens*, Hans Reitzels Forlag, Copenhagen.

McGillivray, D. and McPherson, G. (2014), "Capturing the cultural olympiad in Scotland: the case for multi-criteria analysis", *Cultural Trends*, Vol. 23 No. 1, pp. 18-28.

Monitoring and Advisory Panel (2014), "Aarhus 2017. First monitoring meeting", Brussels, European Commission, available at: https://culture.ec.europa.eu/sites/default/files/files/ecoc-2017-aarhus-monitoring_en.pdf

Myndigheten för kulturanalys (2012), *Att utveckla Indikatorer För Utvärdering Av Kulturpolitik. Redovisning av Ett Regeringsuppdrag. Rapporten 2/2012*, Myndigheten för kulturanalys, Stockholm.

Palmer/Rae Associates (2004), *European Cities and Capitals of Culture*, Palmer/Rae, Bruxelles.

Palonen, E. (2010), "Multi-level cultural policy and Politics of European capitals of culture", *Nordisk kulturpolitisk Tidsskrift*, No. 1, pp. 87-108.

Parkinson, M. and Bianchini, F. (1994), *Cultural Policy and Urban Regeneration : The West European Experience*, Manchester University Press.

Reeves, M. (2002), *Measuring the Economic and Social Impacts of the Arts: A Review*, Arts Council of England, London.

Simonsen (2018), "Welcome future - foreløbige effekter af Europæisk Kulturhovedstad Aarhus 2017", Aarhus: Aarhus 2017, available at: http://www.aarhus2017.dk/da/welcome-future/publikation/index.html

Turku 2011 Foundation (2012), *EUROPEAN CAPITAL of CULTURE TURKU 2011. Final Report of the Turku 2011 Foundation about the Realisation of the Capital of Culture Year*, Turku.

Van den Hoogen, Q. (2012), *Effectief Cultuurbeleid*, Leren van Evalueren, Amsterdam: Boekmanstichting.

Zentrum für Kulturforschung und ICG Culturplan (2011), *Mit Kultur zur Metropole? Evaluation der Kulturhauptstadt Europas RUHR 2010*. Bonn/Berlin.

Further reading

Boland, P., Murtagh, B. and Shirlow, P. (2016), "Fashioning a city of culture: 'life and place changing' or '12 month party'?", *International Journal of Cultural Policy*, Vol. 25 No. 2, pp. 246-265.

Siggins, L. (2021), "Critics happy for state audit to break down Galway 2020 culture funding", Times, May 12, available at: https://www.thetimes.co.uk/article/critics-happy-for-state-audit-to-break-down-galway-2020-culture-funding-qmsn9z78m

About the authors
Hans-Peter Degn is director of Centre for Cultural Evaluation and a teaching associate professor at Aarhus University. He was the evaluation manager of *Aarhus 2017*, in relation to which he developed the initial version of the "Evaluation Dilemmas Model" that this article builds upon. He has been teaching various courses on methods and evaluation for a couple of decades, and during the latter years Hans-Peter has been engaged with the development of methodological approaches within research and evaluation of culture and communication, including development of indicators on, e.g. motivations for cultural participation.

Steven Hadley is an academic, consultant and researcher working internationally in arts management, cultural policy and audience engagement. He currently holds research posts at King's College London and National University of Ireland, Galway and is a Visiting Lecturer at Leuphana University of Lüneburg (Germany) and University College Cork (Ireland). He previously held research posts at the universities of Sheffield, Bradford and Leeds. Steven is an Associate Consultant with both The Audience Agency and Counterculture and sits on the Steering Committee of the Cultural Research

Network and the Editorial Boards of both *Arts and the Market* and the *European Journal of Cultural Management and Policy*. He is Policy and Reviews Editor for *Cultural Trends*. Steven has lectured, taught and delivered training in over 20 countries worldwide and works as a consultant for a wide range of cultural organisations. His recent published work has focussed on cultural democracy and audience engagement. His new book, *Audience Development and Cultural Policy*, is published by Palgrave MacMillan. @mancinbelfast https://www.palgrave.com/gb/book/9783030629694 @mancinbelfast Steven Hadley is the corresponding author and can be contacted at: hadleys@tcd.ie

Louise Ejgod Hansen is Associate Professor at School of Communications and Culture, Aarhus University and Head of Research, Center for Cultural Evaluation. She was project and research manager of rethinkIMPACTS 2017, conducting the research-based evaluation of European Capital of Culture Aarhus 2017. Her main research area is cultural policy, including impacts of culture, evaluation, audience development and participation. She is editor of *Nordic Journal of Cultural Policy* and takes part in different independent and commissioned research projects.