Journal of Educational Change
https://doi.org/10.1007/s10833-023-09480-y

Docility and dilemmas: Mapping 'performative evaluation' and informal learning

Andrew Clapham[1]

Accepted: 8 February 2023

Abstract

Educators working in museums, zoos, and botanic gardens are increasingly required to demonstrate impact. These requirements position 'performative evaluation' as the dominant model, one which also acts as a political, non-neutral, and managerial form of accountability. In contrast, 'practice evaluation' is intended to be democratic, dialogic, and developmental. To explore this contrast, Foucault's concept of the docile body is directed toward interviews with five educators from Italy, Portugal, and the United Kingdom who worked in museums, zoos or botanic gardens. In addition to their work mediating informal learning, all five also had responsibilities to provide evaluation reports to audiences including managers, trustees, funders, policy makers, and politicians. Analysis of these interviews identified a set of dilemmas that the participants faced—dilemmas which illustrate how performative evaluation becomes a disciplinary mechanism which produces docile bodies. I argue that such evaluation is not only inappropriate for the context of informal learning, but undemocratic and non-dialogic. The paper concludes that a reset of performative evaluation from an accountability technology, to a developmental one—along a more sophisticated reading of how informal learning is defined—would not only generate rich evaluate data but mitigate against educators being rendered docile by the process.

Keywords Informal learning · Evaluation · Docility · Dilemmas · Accountability

Introduction

Jenny Ozga (2020) has argued that education policy increasingly installs managerial-technical forms of accountability. Since 2020, there have been seismic geopolitical shifts resulting from economic uncertainty, pandemic, and war—with the

✉ Andrew Clapham
andrew.clapham@ntu.ac.uk

[1] Nottingham Institute of Education, Nottingham Trent University, Room: 321 Ada Byron King Building, Clifton Lane NG11 8NS, UK

Published online: 07 April 2023

fields of education and learning not immune from these shifts. As Collet-Sabé and Ball (2022) note, many of associated discourses have focused on returning education to 'normality'. However, in contrast to those wishing to see such a return, these authors contend that there is an urgent need to 'think education without school…and to start somewhere else' (Collet-Sabé & Ball, 2022, 3).

This paper is set in 'not school' (Sefton-Green, 2012) and focuses on educators' experiences as evaluators and as facilitators of informal learning. Their data not only highlight significant ideas around how and why evaluate, and how we define learning, but also suggests that now might be the time to resist the 'erosion or suppression of democratic[1] possibilities' (Ozga, 2020, 21) resulting from managerial-technical accountability and to '…think differently about education' (Collet-Sabé & Ball, 2022, 2).

The paper is based on interviews with five participants from Italy, Portugal, and the United Kingdom.[2] All five worked in museums, zoos, or botanic gardens as educators (and in one case also as a senior manager) mediating formal, non-formal, and informal learning. They all also had responsibilities to undertake the evaluation of the education provision offered by their institutions and to provide evaluation reports to audiences including managers, trustees, funders, policy makers, and politicians. While there are large scale evaluations of informal learning (for example the Science Learning + project), literature exploring educators' work as informal learning evaluators is relatively small. This paper ventures into this area by providing a lens upon how and why these educators evaluated and the dilemmas they faced in doing so.

Michel Foucault's (1991) concept of the docile subject was mobilised as a theoretical framework (see also Clapham, 2016; Beattie, 2020). Employing this framework highlights how the increase of accountability in public and cultural life (Bulaitis, 2020) has resulted in evaluation becoming a highly 'political'[3] disciplinary mechanism, which renders bodies docile. For the participants, there were two competing drivers for undertaking evaluation. 'Practice evaluation' valued the power of democratic and dialogic evaluation approaches and focused on development (see de St Croix, 2020), in contrast, 'performative evaluation'—theoretically located in Jean-François Lyotard's (1984) work on performativity—focussed/focused accountability.

The participants' dilemmas emanating from these competing drivers not only signalled the challenges that performative evaluation posed them, but also illustrated the complexity of informal learning as a concept. For them, informal learning could as much occur via structured (even though non-curricular) learning activities in museums, zoos, and botanic gardens as via unstructured learning activities occurring in a café or art gallery. For example, a 'guided walk' could mediate learning informally and visitors taking a class in flower illustration might also informally learn about the biological functions of flowers.

[1] Here, 'democratic' is a means of challenging hierarchical power relations (Pateman 2012).

[2] All names used in the paper are pseudonyms (see BERA, 2018).

[3] Politics is 'power acting on power' (Sluga, 2011).

Table 1 Learning typology (Amended from Eshach, 2007, 174)

Formal learning	Non-formal learning	Informal learning
Usually at school	At institution out of school	Everywhere
May be repressive	Usually supportive	Supportive
Structured	Structured	Unstructured
Usually prearranged	Usually prearranged	Spontaneous
Motivation is typically more extrinsic	Motivation may be extrinsic but it is typically more intrinsic	Motivation is mainly intrinsic
Compulsory	Usually voluntary	Voluntary
Teacher-led	May be guide or teacher-led	Usually learner-led
Learning is evaluated	Learning is usually not evaluated	Learning is not evaluated
Sequential	Typically non-sequential	Non-sequential

Rather than attempting to make hard delineations between learning types, the participants described informal learning as mediated by a complex intersection of experiences, discourses, and interactions rather than by where it might occur—and consequently that evaluating such learning requires a far more sophisticated and nuanced approach than simply focusing upon impact.

These definitional inconsistencies, in conjunction with the pressures of performative evaluation, posed significant dilemmas for these educators as evaluators. I argue that these dilemmas highlight fundamental questions not only about evaluation but how learning, knowledge, and education are considered. Like authors such as de St Croix (2020) and Coultas (2020), the paper offers a counter to performative evaluation by suggesting that a non-disciplinary, democratic, and dialogic model would shift evaluation beyond a narrow focus upon impact and accountability. Not only would such a model provide rich evaluative data but would reset evaluation from a disciplinary technology and mitigate against educators being rendered docile by the process.

Informal learning

Learning that takes place in school is often defined as formal and that which occurs outside school is informal. Gerber et al., (2001, 570) for example, suggest that informal learning is the 'sum of activities' when learners 'are not in the formal classroom in the presence of a teacher'. With the prominence of such definitions, perhaps it is unsurprising that much of the work exploring informal learning is set outside school and within the workplace (for example Eraut, 2004). Those who have ventured beyond the workplace to explore informal learning however (Clapham, 2016; de St Croix, 2020; Falk, 2001; Falk & Dierking, 2010, 2018; Howard, 2021, 2022; Sims, 2019), suggest it plays a significant role in the way children, young people, and adults learn.

Nonetheless, the term 'informal learning' is not uncontested and the use of 'formal', 'non-formal' and 'informal' in relation to learning are often conflated (Quinn,

2018). For Eraut (2004), definitional inconsistencies around informal learning are exacerbated by typologies (for example Table 1) which often employ the settings and architecture where learning takes place, and how learning is organised, to delineate between learning types.

Such typologies often describe informal learning as occurring 'everywhere' as well as being 'supportive; unstructured; spontaneous; intrinsically motivated; voluntary; (usually) learner-led; not evaluated and non-sequential' (Eshach, 2007, 174). However, describing informal learning as occurring 'everywhere' appears to contradict the use of school and not-school as way of defining what type of learning is taking place. Clearly, if informal learning occurs everywhere then it is just as likely to take place in school as in not-school.

Contradictions such as this add to the debate around the efficacy of defining learning in schools as formal, and that occurring elsewhere as non-formal (Rogoff et al, 2016). The 'slipperiness' of informal learning as a concept is exemplified in several other ways. For example, many museums, zoos, and botanic gardens have separate buildings, such as 'learning centres', which replicate a school classroom (see Cunningham & Walton, 2016). As well as having tables and desks set out as if in a school, learning in these spaces can be organised via a curriculum, schemes of work, and learning outcomes and have assessments (all of which suggest it is a formal learning setting). Nevertheless, both non-formal and informal learning activities can also take place in these buildings (Berman, 2020).

Authors such as Maarschalk (1988), Sim (2019) and Jamison (2009, 2013) acknowledge the complexity of informal learning as a concept and provide nuanced and sophisticated descriptions of it. Jamieson (2013, 145) for example describes it as:

> …a complex web of experiences and interactions, undertaken over a wide range of physical environments, from internal to external spaces, including classrooms, cafes, plazas and libraries.

With this complexity in mind, Jamieson's ideas around informal learning are adopted here. Whist Jamieson also uses architecture as a means of understanding where and when informal learning occurs, he stresses it can take place as much in a classroom and a library as in a café or plaza. Considering informal learning as complex, experiential, interactional, and not restricted by location or architecture highlights how making hard delineations between learning types run the risk of ignoring this complexity, as well as posing significant challenges for those evaluating such learning.

Evaluation

Initially, defining evaluation appears unproblematic as there are well established definitions accepted across many disciplines (Table 2).

However, these definitions present challenges for those evaluating informal learning, as they suggest that evaluation broadly considers feasibility, impact, value for money, or processes. When considering the complexity of informal learning, Jeffs

Table 2 Evaluation typology (Amended from NFER, 2022)

Pilot evaluations and feasibility studies	Appropriate for evaluating initiatives in the early stages of development and for assessing whether a larger-scale evaluation is sensible and viable
Impact evaluations	Investigates the impact of an initiative, drilling down into the effect of the intervention over and above what might have happened without it and the extent to which any changes can be directly attributed to the intervention. Approaches such as randomised controlled trials (RCTs), quasi-experimental and pre- and post-intervention designs are employed
Value for money evaluations	Assess whether an intervention is delivering value for money, compared to similar initiatives, as part of a wider impact evaluation
Process evaluations	Explores how well an initiative was implemented and which use a range of qualitative and quantitative methods

and Smith (1999) and de St Croix (2020) argue that evaluation models focusing on impact and accountability are inappropriate for such learning. Moreover, the 'gold standard' status afforded to highly impact focused evaluations—which employ for example randomised controlled trials (RCTs)[4]—are not uncontentious, with concerns being raised regarding their appropriates and efficacy when applied to education and learning (Dawson, et al., 2018; Styles & Torgerson, 2018).

An alternative way of considering evaluation is as a political technology (Banner, 1974; Wergin, 1976). There is a significant body of work highlighting how accountability is increasingly central to education policy (Lingard et al, 2013) and arguing that evaluation is non-neutral (Weiss, 1993), 'imbued with power' (Taylor & Balloch, 2005, 1), and highly political (Eckhard & Jankauskas, 2020; Vestman, & Conner, 2006). Describing evaluation in this way provides a nuanced means of understanding its role in internal and external disciplinary matrices and how these matrices have become ubiquitous in contemporary organisational practices.

This ubiquity has made evaluation a 'booming industry' (Eckhard & Jankauskas, 2020, 695) which brings into focus considerations around 'market pressure' and the relationship between evaluator and 'customer'. Van Voorst and Mastenbroek, (2019) for example, argue that the demands made by the customer for optimal evaluation findings can result in significant pressure exerted on the evaluator, whilst Eckhard and Jankauskas (2020, 685) warn that that the '…political use of evaluation may hinder its functional purposes'.

Considering evaluation as non-neutral and political therefore moves its analysis beyond simply considering 'how we evaluate' and toward a far more nuanced and

[4] See Picciotto, 2014; Camfield and Duvendack, 2014 who critique RCTs as the 'gold standard' methodology for evaluation research.

sophisticated reading of 'why we evaluate'. As I go on to discuss, the participants' data illustrates how performative and political pressures shaped the 'functional purpose' of the evaluations they udertook. Moreover, the dilemmas that performative evaluation posed them illustrate not only about the pragmatics of 'doing' evaluation, but also offer a glimpse of ways in which established models might be challenged.

I now explain how Foucault's (1991) concept of the docile body can be employed as theoretical framework.

Theoretical framework

The focal point for Foucault's (1980, 1991, 2000, 2010) thesis around the production of docile bodies is disciplinary power. The materiality of bodies (both biological and institutional) are the sites upon which power is produced and exercised (see also Dreyfus and Rabinow, 1983). For Foucault power is not a duality of being either oppressive or liberatory, nor is it 'wielded' by one upon another. Rather, it is relational and plays out through a range of disciplinary matrices (Ball, 2013; Foucault, 2000; Rabinow & Rose, 2003) which are mediated via discourses, practices, and institutions. Disciplinary power is also underpinned by 'rules of conduct' (Foucault, 1991), with the fear of being observed breaking these rules, resulting in subjects disciplining themselves as much as being disciplined by others (see Margolis & Fram, 2007).

The intersection between disciplinary power and the rules of conduct ultimately lead to the production of the docile body. Bodies become docile when they 'subjected, used, transformed and improved' (Foucault, 1991, 136), via disciplinary matrices which define what is 'normal' and 'deviant' behaviour (Foucault, 2010). For Foucault (1991) docile bodies are produced through the 'art of distributions' (141) and via:

1. *Enclosure* (Foucault, 1991, 141–142) enables bodies to be grouped together through architectural and organisational processes and acts as a means of regulation. Enclosure can be achieved via physical actions, such as confinement, but also via architectural and organisational technologies such as barracks, factories, and monasteries.
2. *Partitioning* (Foucault, 1991, 143) enables enclosed bodies to be systematically distributed within groups to 'establish presences and absences …to be able at each moment to supervise the conduct of each individual, to assess it, to judge it, to calculate its qualities or merits' (1991, 143). Partitioning is a sophisticated means of grouping bodies and extends the capabilities of enclosure.
3. *Functional sites* (Foucault, 1991, 143–144) are 'coded spaces' which enable individual bodies to be analysed. Foucault (1991, 144) uses a naval hospital, in a naval port, as an example of such a site and maps how the disciplinary effects of the hospital play out across every aspect of the port. Coded spaces are so effective in mediating disciplinary power that their effect can be extrapolated to entire labour process.

4. *Ranking* enables individual bodies to be compared and to be circulated in a 'network of relations' (Foucault, 1991, 145–146), which can be at once 'architectural, functional and hierarchical' (148). Ranking is so fundamental to disciplinary systems, and in producing docile bodies, that Foucault (146) describes discipline is an 'art of the rank'.

I want to particularly draw attention to the ways in which these aspects of disciplinary power can be applied to performative evaluation and how it works as a 'disciplinary tactic' (Foucault, 1991, 149). Whilst such tactics illustrate how evaluation produces docility in the evaluator and evaluated, they also frame it as non-neutral, political, undemocratic, and non-dialogic. Narrating performative evaluation in this way illustrates the complex relations and interactions that make up disciplinary systems and how such evaluation becomes far more than simply a tool for understanding 'how something works'. The participants' data illustrate how such evaluation operates as a highly effective form of managerial-technical accountability which renders bodies '…manipulated, shaped, trained, which obeys, responds' (Foucault, 1991, 136) and ultimately both disciplined and docile.

In the following section I outline the methodology and analysis employed and which resulted in the two themes, and associated dilemmas, which are the focus here.

Methods

The paper reports on a qualitative study that employed naturalistic semi-structured interviews (see Kvale & Brinkman, 2009) as the primary data generation method. In total 19 interviews with the five participants took place (Tanya $n=3$; Luca $n=5$; Sophie $n=4$; Max $n=2$; Alessandro $n=6$)—see Table 3.

All five participants were part of a larger cohort of 79 educators—Italy ($n=31$); Portugal ($n=23$); and the UK ($n=25$)—participating in the BGCI/Erasmus + (2019) LearnToEngage (LTE) project. Facilitated as a blended learning course, LtE developed a suite of professional development modules for zoo, botanic garden, and museum educators aimed at enhancing engagement with audiences and supporting the educational role played by these institutions. Alongside the Interpretation, Working with Diverse Audiences and Science Communication modules, Evaluation was one of the LtE modules offered to participants and which the author co-led.

As part of the LTE project, I undertook interviews with 39 of the participants regarding their experiences as evaluators. Of these, five offered to take part in a further set of interviews specifically about evaluating informal learning which form the basis of this paper. These interview data were generated during coffee breaks or lunchtimes or during social events in the evening.[5] Consequently, the participants comments were captured as reflexive fieldnotes, which are not verbatim accounts of what was said, but my understanding which I later confirmed for accuracy with the participants (see also Clapham and Vickers, 2018).

[5] The project received favorable ethical opinion from the university Ethics Committee and all participants provided ethical consent (see BERA, 2018).

Table 3 Informant biographies

Educator	Biography
Luca	Educator based in northern Italy and who works extensively facilitating informal learning with school, university, and many other visitor groups as part of an Education Department consisting of 10 academics. Luca has a master's degree in science communication and has a particular focus on informal learning and science communication. He has been an educator for 6 years and has been part of both designing or implementing evaluation programmes in this time.
Tanya	Educator with 10 years' experience and has a PhD and has been a secondary level science teacher for 6 years. Tanya has led her garden's informal learning around interpretation and its work with diverse audiences of all ages. Tanya has evaluated informal learning but has not seen recommendations implemented.
Sophie	Educator with 18 years' experience of working in a small setting in the UK. Sophie described her work as curator, strategy planner, teacher, and cleaner all wrapped up in one. Sophie has a teaching qualification and was a secondary level teacher for 5 years. Sophie has a master's degree in education and has particular interest in informal learning and with adult learners from vulnerable groups. Sophie designs and implements all the informal learning evaluation that takes place.
Max	Director of a small organisation in the north of the UK that has charitable status but also receives, and bids for, public funding. Max has worked in various roles in the botanic garden and museum sector for 10 years and has a master's degree in business administration. He uses external consultants to design and implement evaluations.
Alessandro	Educator in his first year working in a setting in Italy—holds a PhD in science communication. Has a particular interest in informal learning and young people in 18–24 age group from challenging socio-economic backgrounds. He has had limited experience of designing an undertaking informal learning evaluation in his current role, although this formed an extensive part of his PhD thesis.

Employing the participant's data in this way is not about establishing definitive meaning. Rather as Allard and Doecke (2017, 285) suggest, such data can be read as 'symptomatic of a larger policy landscape' and provide an opportunity to theorise around that landscape. I employed Braun and Clarke's (2020) reflexive thematic analysis to explore these data. Drawing on their approach, I familiarised myself with the data, undertook initial code generation, and sorted those codes into themes. I then added an additional layer of coding by focusing on the participants' 'dilemmas' and how these mapped to, and across, the two themes.

In the following sections two main themes—'doing evaluation' and 'why evaluate'—along with some of the participants' dilemmas associated with them, are discussed.

'Doing evaluation'

The first theme 'doing evaluation' had four key dilemmas related to it: evaluation as game playing; how evaluation was resourced; what methods and data were considered 'gold standard'; and what appeared resistance to democratic and dialogic alternatives.

Central to all these dilemmas was how evaluation had become predominantly an accountability exercise required by managers and trustees (internally) and funders, policy makers, and politicians (externally). For all five participants there were significant implications of this, with all of them using the same metaphor around 'game playing'. For them, one of the key requirements of playing the game was to commoditise and monetise informal learning (see also de St Croix, 2020), with impact focussed evaluation taking on a key role in this.

As Sophie outlined, successfully negotiating the economic reality facing museums, zoos, and botanic gardens required not just a will (for the participants and their intuitions) to play the game but also a highly "strategic approach" to doing so. For Sophie, there were significant consequences resulting from the requirements to play the game, with one of the most prominent being the "impoverishment" (Max) of that which was considered gold standard evaluation evidence. Luca meanwhile was highly conflicted when it came to making causal claims around the impact of informal learning:

> …we're supposed to show that this amount of informal learning equals this amount of something else…but how? How can I say that this piece of interpretation, means that this much [informal] learning happens, which has this impact on the visitor, which has this impact on the economy? It's about playing the game though…

The dilemmas associated with playing the game were illustrated in several ways. For example, all five participants resented how the impact agenda, and the number of resources directed toward it, had reduced their capacity to do work in other areas. This resentment was only heightened as they considered impact focussed evaluation as methodologically flawed when directed toward informal learning.

For Sophie evidencing the impact of informal learning via hard data that was "tenuous at best", whilst Max felt that the hunt for impact, had reduced informal learning to a set of easily quantified proxies such as visitor numbers, attendance and engagement (see also Joslin, 2021). One example of this was the way impact focused evaluation was considered more legitimate than more practice focused alternatives. The participants' dilemmas around this methodological legitimacy were further exacerbated by their research backgrounds as two of them held PhDs in museum education whilst the others had Master level degrees in the social sciences. Their background as researchers led to them question the efficacy of making causal claims concerning the impact of informal learning:

> I feel very uneasy about relying on proxies as measures [of informal learning] and then using them to make causal claims…it's not that simple. (Alessandro)

The other participants also questioned the efficacy of impact focused evaluation of informal learning. They felt the high status afforded to statistical analysis of large scale data sets using RCTs was particularly misplaced for informal (and non-formal) learning. Like Picciotto, (2014) and Camfield and Duvendack (2014), the participants recognised the power of RCTs, but also stressed that this approach was not without methodological challenges.

This notion of 'gold standard' evaluative evidence and methodologies was problematic for all five participants. Max for example outlined the pressures to provide "bottom line" evidence and how these pressures posed substantial dilemmas for him. Although he had made significant efforts to promote the use of practice evacuation, his trustees were resistant to it. Similarly, the trustees rejected his concerns around the efficacy of using proxies such as ticket sales and attendance to illustrate the impact of both non-formal and informal learning. Consequently, he was "fighting a losing battle" to persuade trustees to support the use of alternative evaluation approaches. Not only did the trustees consider performative evaluation as the 'industry standard', but these models were relatively easy and cheap to employ. Moreover, the trustees questioned what alternative evaluation approaches could offer over impact focused evaluation and were concerned how such alternatives would be received by funders, policy makers, and politicians.

The participants' dilemmas resulting from this narrow view of evaluation methodologies were even more frustrating as all five participants discussed the availability of alternatives (see also Allen & Peterman, 2019). As Sophie outlined, museums, zoos, and botanic gardens had a history of developing and employing innovative methodological approaches toward generating evaluation data:

> There're innovative ways of working with learners to understand what informal learning means for them. That's what's so frustrating…there're tools and methods we could use and we're still counting tickets.

Despite the availability of alternative evaluation approaches, the participants reported these were seldom if ever utilised in their institutions. In their experience, managers, trustees, funders, policy makers, and politicians rejected non-performative evaluation approaches as being too time- and resources-intensive, or simply disregarded as not 'proper' evaluation at all.

Why evaluate

These dilemmas around 'doing' performative evaluation were interconnected with those regarding 'why' such evaluation should be undertaken. Regarding this theme, the participants had five further dilemmas: the need to be pragmatic evaluators; the purpose of education and being an educator; being a bystander; challenging beliefs; and resetting (and resisting) accountability.

Regarding 'why' they evaluated, Tanya's comments encapsulated the feelings of all five participants. For Tanya, the way that evaluation was primarily part of disciplinary and accountability processes, rather than being development, resulted in her 'unease':

> I feel really uneasy about evaluating [learning] when it's all about accountability…it's not what evaluation should be about.

Tanya's comments were reflected by the other participants, who also reported that rather than being a form of accountability, evaluation should be

developmental. Despite this unease, all five were pragmatic as to why they evaluated, and Tanya again used the game playing metaphor:

> [why do you evaluate…] To play the game! If we don't play the game then we don't get funding...and it's pretty obvious what that means…I don't have a choice.

Like Tanya, the other participants also acknowledged that undertaking performative evaluation was part of playing the game, and they were pragmatic about playing it. However, as much as their pragmatism was about playing the game, it also reflected the accountability culture facing them and their institutions. Consequently, the demands of performative evaluation, and the requirement to play the game because of them, had a twin effect. Not only did these render the participants docile, but also firmly categorised them as examples of 'performative workers' (Ball, 2003).

Tanya and the other participants were starkly aware that being such a worker meant that they were being enclosed, partitioned, and ranked. However, all five felt that the ubiquity of performative systems and technologies meant there was no alternative but to play the game and be ranked. Nonetheless, there was admiration for those educators and institutions who had successfully played the game, with Luca citing as an example economic impact evaluations (see for example Trainer, 2010):

> I see the way some [museums] use economic impact and it's clever. But we don't have the funds to pay for an econometric analysis of our impact and besides, how can you draw causality between learning whilst sitting on a park bench and economic impact…

This pragmatism, however, illustrated the participants' dilemmas not only about evaluating informal learning, but around macro scale discourses concerning the purpose of education. For Tanya, the demands made upon her and her institution to be accountable meant she increasingly felt like a "bystander", not only as an evaluator, but in relation to the trajectory of how education and learning were considered more widely:

> I don't feel that I'm doing evaluation, or I have a say in what [informal] learning is, I'm just part of a process…a bystander...

This notion of 'bystander' was echoed by the other participants. They increasingly felt disenfranchised by the tensions between, on one hand, the demands for impact, and on the other, their beliefs around the core purpose of their work as educators and of museums, zoos and botanic gardens as educational intuitions. Performative pressures, practices, and cultures—of which evaluation was just one—meant they were all considering career changes (see also, Rende et al, 2021). All five felt that although they loved their work, the "sacrifices" (Sophie) that being an educator involved (low salary, lack of career progression, low standing of educators and education departments within their institution) increasingly outweighed their job satisfaction.

These dilemmas around how game playing was changing the purpose of their work, and why they were educators, exposed the participants' deep seated beliefs

around far more than just evaluation. For all five, the accountability culture they and their institutions increasingly inhabited was diametrically opposed to the philanthropic and social drivers which led to their establishment. For example, Sophie outlined how the educational benefits mediated by museums, zoos, and botanic gardens—as well as their environmental, social, emotional, and health benefits (see, Jordan, 1994)—were central to why many of them were originally built:

> When they were built…public parks and gardens were informal learning spaces as much as for leisure activities…and they were linked to museums and galleries…learning was part of them.

All five participants mapped out this disconnection between the philanthropic and social values underpinning the initial construction of museums, zoos, and botanic gardens and the current accountability environment. For Alessandro, this disconnection was so central that it signalled an "existential threat" to his work as an educator and to that of his institution as a place of learning.

Similarly, Luca recounted how the threats posed by accountability were far wider reaching than simply for him, his Department, and his museum. For Luca, if museums, zoos, and botanic gardens were to continue to mediate learning of all types then a reset and broadening (see Harrison, 2014) of accountability structures was required:

> I need to be accountable, but it works just one way. What about my workplace being accountable to me…what about funders and government being accountable for how much they help us to do our jobs?

Like Luca, all the other participants also felt that they should be held accountable to managers, funders, trustees, and most crucially the public. However, they also highlighted how accountability appeared to be 'one-way', as they (and their institutions) were held accountable, but trustees, funders, policy makers, and politicians appeared to be less so. All five felt that accountability should be 'two-way', so that they could hold others—managers, trustees, funders, policy makers, and politicians—accountable for the extent to which they supported these educators to undertake their work (see also Gewirtz & Cribb, 2020).

Despite advocating such a reset, all five were gloomy as to the likelihood it would occur. For them, without a fundamental reorientation of how accountability was considered, it was unlikely that such a reset would gain traction in education systems locally or nationally let alone globally.

Discussion

What we see from the participants' data was the extent to which disciplinary, managerial-technical accountability was directing how and why they evaluated. We also see how such accountability rendered them and their institutions both docile bodies and performative workers.

Many of their dilemmas were concerned with how evaluation was increasingly an accountability and game playing exercise. For them, evaluation as game playing was evident at the micro scale of day-to-day practices and processes (of which evaluation was one), and at the macro scale of policy discourses concerning the purpose and structure of education and learning that went beyond informal learning.

The game playing metaphor was illustrated in several dilemmas. For example, if the participants 'played the game'—by proposing to undertake impact and accountability focused evaluation—then not only were such evaluations likely to be resourced, but upon completion the participants would be congratulated by managers, trustees, funders, policy makers, and politicians on a 'job well done'. In contrast, if they proposed to undertake developmental, dialogic, and democratic evaluations, these were unlikely to be funded as they were considered (again by managers, trustees, funders, policy makers, and politicians) as lacking the 'hard' evaluative data required.

This 'choice'—evaluate performatively and be rewarded or evaluate developmentally, dialogically, and democratically and be punished—also related to those dilemmas concerning the methodological efficacy of performative evaluation. For the participants, the demands for 'hard' data had reduced informal learning to a set of easily quantified proxies. Whilst such proxies presented the opportunity to generate such data, in their view, they fell well short of representing the richness and complexity of informal learning.

This narrow view of evaluation methods and data was even more exasperating for the participants as there were alternatives. They recounted how museums, zoos and botanic gardens had a track record of developing alternative and innovative approaches toward generating evaluation data such as gaze eye tracking (Dondi, et al, 2022) and visitor to visitor learning (Pitts, 2018). Despite these alternatives, the participants saw issues of resource and methodological legitimacy as central as to why such alternative evaluation methodologies were not adopted more widely.

The participants' dilemmas around 'doing' performative evaluation, intersected with those regarding why undertake such evaluation at all. For all five there was a simple answer to this question: they could either play the game or resist it. This pragmaticism to play the game was not without cost though, as it resulted in them experiencing inner conflicts and inauthenticity. The options to play or resist were not a duality however, but rather a continuum. At some times they played the game, at others they resisted it.

This ebb and flow in the participants' resistance resonated with Foucault's (1991, 2000) ideas around 'day-to-day resistance'. Indeed, much the participants' resistance echoed Foucault's analysis that resistance is inherent within relations of power (see also Ball, 2013). Nonetheless, the sheer scope and effectiveness of disciplinary technologies such as performative evaluation—and despite the participants' day-to-day resistance—meant that they were still rendered docile by them. For the participants, therefore, this intersection between discipline and docility reflected fundamental, macro scale discourses concerning the purpose of education. Again, the effectiveness of disciplinary power was illustrated in how the participants were rendered docile 'bystanders' in not only the evaluation process, but also in much of what education systems were becoming.

The participants' feeling of being a bystander also reflected their dilemmas concerning the purpose of their work, their careers, and of museums, zoos, and botanic gardens as places of learning more generally. They saw a disconnection between the philanthropic drivers that led to these institutions to be originally built and the contemporary performative landscape they inhabited. Moreover, this conflict between the participants' beliefs and the need to evidence impact was reflected in dilemmas around their careers. All five participants were increasingly considering career changes as the pressures of accountability brought into question if the sacrifices required to be an educator (low salary, lack of promotion prospects, the low standing their work and that of education departments within their institutions) were worthwhile.

The participants acknowledged that their dilemmas were not solely a consequence of accountability structures; indeed, all five felt evaluation was legitimate requirement of their work. However, performative evaluation was just one more element of an increasingly impact focused workplace. For them, and like Ozga (2009), the drive to demonstrate impact reflected how museums, zoos, and botanic gardens and education systems more widely, were increasingly 'data driven and data governed'.

Performative evaluation and the docile body

The participants' data highlights how performative evaluation worked as an accountability technology producing both docile bodies and performative workers. For Foucault (1991, 136) docility was primarily concerned with joining the 'analysable body to the manipulable body'. When we consider performative evaluation as a means of achieving this, not only do we see it as a powerful means of producing docility but also how it is non-neutral, highly political, undemocratic, and non-dialogic.

The sophistication of performative evaluation as a means of producing docile bodies is perhaps best reflected by the role the bodies themselves take in this process. Foucault (1991, 170) suggests that discipline 'makes' individuals and that is it 'the specific technique of power that regards individuals as both objects and instruments of its exercise'. The participants' data suggests that they were patently aware that performative evaluation was a disciplinary technology which had the aim of making them an efficient and docile body.

The way performative evaluation 'made' the participants docile was particularly evident in how it both enclosed and partitioned them (as biological bodies), as well as the institutional bodies where they worked. Such evaluation enabled bodies to be enclosed and grouped together so that they could be compared with one another, as well as to keep 'order and discipline' (Foucault, 1991, 142). Evaluating against internal and external targets reflected much of Foucault's description of how discipline defines what is 'normal' and 'abnormal' behavior and the divisions between these behaviours. Consequently, performative evaluation was both the means for how the participants were enclosed and partitioned, as well as acted as the rules against which their behaviour was judged and ranked.

As Foucault (1991, 143) notes, 'disciplinary space tends to be divided into as many sections as there are bodies' and performative evaluation is highly effective in

mediating this division, as enables the grouping of multiple bodies together, in multiple disciplinary spaces. The way performative evaluation enables this division partitions and systematically distributes bodies with the outcome that the 'meticulous control and operations of the body' (Foucault, 1991, 137) was possible.

Perhaps the most effective part of this partitioning was how the participants' own actions of 'playing' the accountability 'game' partitioned them as much as the actions of others. Nonetheless, their pragmatism in choosing to 'play the game' was perhaps entirely understandable as they had no choice but to comply with the 'normalizing judgments (Foucault, 1991, 177) inherent in performative structures. Resisting macro-scale structures such as performativity was beyond the capabilities of these five individuals and consequently, they made the pragmatic choice to play the game, and to be both enclosed and partitioned as a result.

What this pragmatism also meant was that performative evaluation acted as a functional site. The demands from management, funders, and government for performative evaluation to demonstrate impact meant it not only partitioned bodies but acted as a 'coded space'. Such spaces play a crucial part in in enabling the 'presence and application' (Foucault, 1991, 145) of bodies to be observable and analysable—analysis which in turn provided the evidence as to whether those bodies require disciplining or rewarding.

The fixation upon demonstrating impact meant that as much as performative evaluation was highly effective as a coded space, it was similarly effective as a means of ranking. The participants' data suggests that the way performative evaluation mediates ranking is one of its most powerful attributes as a disciplinary technology. What this attribute also means is that it becomes ubiquitous within contemporary institutional processes and the 'hierarchical observations' (Foucault, 1991, 170) it mediates are accepted as the norm. This acceptance also shows the way performative evaluation acts as one of what Foucault (1991, 170) calls the 'simple instruments' of disciplinary power: 'hierarchical observations, normalizing judgment and their combination'.

The way evaluation works as hierarchical observation and normalizing judgment, and how it is also used to rank bodies, has two outcomes—it renders bodies docile and ensures that individuals are 'good' performative workers (Ball, 2003). For Ball (2003, 215), contemporary organisational and work practices mean that the performative worker is forced to set aside '…personal beliefs and commitments' and lives '…an existence of calculation'.

What the cases in point suggest, is that the docile subject and the performative worker are one in the same. For Ball (2003, 215):

> The new performative worker is a promiscuous self, an enterprising self, with a passion for excellence. For some, this is an opportunity to make a success of themselves, for others it portends inner conflicts, inauthenticity and resistance.

Moreover, Ball (2000) goes on to describe how performativity has resulted in workers adopting 'cynical compliance' of performative systems. Ball's description of such compliance resonated with the participants' use of the game metaphor. It also mapped to their analysis of the powerful, sophisticated and subtle ways that performative systems, and playing the game, rendered them docile.

The participants felt they had no choice but to play the performativity game, which resulted not only in docility but in them suffering inner conflicts and inauthenticity. Consequently, although performative evaluation ostensibly led to the increased efficiency central to being a performative worker, it also resulted in bodies (educators and institutions) who were disenfranchised, disempowered, and docile. This docility did not come without a very high cost. For the participants, it signaled an existential threat not only to the core purpose of their work as educators, but to museums, zoos, and botanic gardens as educational institutions.

Toward and alternative

Whilst the participant's experiences suggest there needs to be an alternative to performative evaluation, they also highlight the need for a sophisticated re-conceptualisation of how informal learning is defined and understood.

Although museums, zoos, and botanic gardens could be described as non-formal learning settings, the participants argued that informal (and formal) learning also took place in them. Consequently, they defined informal learning as part of a complex intersection of interactions, relationships, and emotions rather than confined to types of architecture or the way learning is organised.

Table 4 captures some examples of activities that took place in the participants' institutions, which they felt could mediate informal learning. This list is not exhaustive and many of these activities could also mediate non-formal learning. However, if the participants felt there were opportunities for informal learning to occur during them, they are included.

What the participants' re-conceptualisation of informal learning also highlights is how for them such learning had to be evaluated democratically and dialogically rather than performatively. The participants are not alone in this conclusion, as de St Croix (2020) also argues the case for dialogic, inclusive, and democratic approaches of evaluating informal learning.

Like de St Croix, the participants proposed that if evaluation was to become more practice focused, then the dominant performative cultural episteme also required essential reorganisation. For them, one way this could occur was by embedding alternative accountability models that are non-disciplinary, democratic, and dialogic within contemporary organisational practices. Table 5 draws on the cases described here to outline the main differences between performative and practice evaluation. Table 6, meanwhile, highlights some of the ways that practice evaluation can be used to evaluate informal learning.

What Table 5 and 6 suggest is that there are alternatives that could lead to a reset of performative evaluation. Although such a reset might appear to be fanciful, the stories recounted here suggest that if evaluation is to be more than an accountability tool, then such a fundamental reset is both possible and essential.

Table 4 Informal learning activities

Informal learning activity	Indicators	Programme examples
Reflecting	Reflective diary; wellbeing journal	Citizen science; Community Learning; Family and Early Years learning
Investigating	Quadrats; science table; trial and error; observing	Citizen science
Moving	Unguided tour; sensory tours; guided unstructured tours; dance; drama	Interpretation, working with diverse audiences
Researching	Reading; smart devices; browsing internet. social media; QR codes; scientific enquiry; plant science	Family nature initiatives
Making	Art; music; creative writing; poetry; tapestry; singing; photography; paper flowers; illustrating; painting; Knitting; carpentry; terrariums; aromatherapy; herbal preparations; weaving	Community learning; family and early years learning
Growing	Gardening workshops; community allotment; local green initiatives; plant selections; garden design; horticulture	Community horticultural learning
Volunteering	Tours; meeting and greeting; gardener/allotment helper;	Community learning; youth; community outreach
Explaining	Citizen science; interpretation	Outreach
Talking	Youth forum; family conversations; dialogue	Family learning
Listening	Audio tours; talks; songs; poems; storytelling	Talks
Playing	Games based activities; music and movement;	Family learning

Table 5 Performative versus practice evaluation

Performative evaluation	Practice evaluation
Driven by metrics	Driven by narratives
Focuses upon impact	Focuses upon experience
Likely to employ mostly quantitative methodologies	Likely to use mixed methods and to value qualitative methodologies
Linked to funding	Linked to practitioners investigating their practice
Given high status	Although ostensibly given high status, actually has low status
Likely to generate large data sets	Likely to generate small data sets
Likely to make direct comparisons	Unlikely to make direct companions
Has an audience of senior leaders, policy makers and funders not users and practitioners	Has an audience of users and practitioners with senior leaders, policy makers and funders less likely to engage with findings

Table 6 Practice evaluation of informal learning

Area of investigation	Informal learning evaluation research tools and strategy
Identifying what leaners know	Self-audit of existing knowledge prior to visit(s)
	Comment banks
Identifying why learners learned	Reflective diary
	Self-audit of learning motivation
Identifying what learners learned	Reflective diary
	Self-audit of knowledge post visit(s)
	Peer-audits sharing what each other learned
	Learning portfolios
	Interviews
	Focus groups
	Learner panels
	Online social networks
	Longitudinal case studies
Identifying how learners learned	Analytics—dwell time
	Self and peer audits of interpretation
	Peer and self-audits of science communication
	Learners' co-construction of interpretation
	Learners' co-construction of science communication
Mobilising what learners tell us about our learning offer	Learner forums
	Leaners contribute to strategy
Disseminating what learners tell us about our learning offer	Learner voice included in dissemination of evaluation findings
	Disseminating findings for different audiences

Conclusion

The cases reported here highlight discourses around why and how learning should be evaluated. However, they also go beyond evaluation by also raising fundamental questions concerning the purpose of education and how learning is defined. They illustrate the way that accountability constrains and restricts evaluative practices and how evaluation acts as a highly effective political, disciplinary, and accountability mechanism that produces docile biological and institutional bodies.

However, whereas Foucault (1991, 136) described 'projects of docility' which place the body 'in the grip of very strict powers' which impose 'constraints, prohibitions or obligations' the participants in this study at least saw an alternative. For them, non-disciplinary, democratic, and dialogic evaluation would enable a reset of evaluative processes and practices from being accountability technologies to developmental ones. Such a reset would not only reframe evaluation, but would mitigate against educators and the museums, zoos, and botanic gardens where they worked being rendered docile as a result.

Such a reset chimes with Garcia's (2012) call to revisit the way museums describe themselves:

> It is time to revisit the way we describe and advocate for the "learning power" of museums…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...

The participants also stressed the importance of mediating the 'intrinsically-motivated, joyful, open-ended learning' Garcia describes. However, they were less inclined to blame themselves, their colleagues, or their institutions for 'not doing enough'. Rather, they felt they were doing their best to play the game, whilst also recognising they were rendered docile in doing so.

The dilemmas described here reflect the omnipresent discourses around quality, standards, and impact—and how performative evaluation mediates these—that are 'front-and-center' across education institutions and systems globally. However, this is not an argument to abandon evaluation, rather, a plea that the focus upon evaluation as a performative, accountability and disciplinary technology needs to be reset. At the outset, I cited Jenny Ozga's (2020) analysis of education policy increasingly installing managerial-technical forms of accountability. With the current geo-political climate in mind, I maintain that the educators' experiences outlined here signal how now might be the ideal time to resist the erosion and suppression of democratic possibilities Ozga describes whilst enabling us to '…think differently about education' (Collet-Sabé & Ball, 2022, 2).

Clearly, such resistance has implications beyond simply evaluation. However, in part at least, re-imagined evaluation models would challenge and reset the top down requirements to evidence impact that Ozga outlines. If the inequalities that

are inherent to education systems are to be challenged, performative accountability needs to be reset and practice evaluation valued. Doing so would mitigate against educators, and the museums, zoos, and botanic gardens where they work, being rendered docile and would contribute to a rethinking of education systems at the local, national, and global scales.

References

Allard, A. C., & Doecke, B. (2017). Telling tales: The value of storytelling for early career teachers. *Pedagogy, Culture and Society, 25*(2), 279–291.

Allen, S., & Peterman, K. (2019). Evaluating informal STEM education: Issues and challenges in context. *New Directions for Evaluation, 2019*(161), 17–33.

Ball, S. J. (2000). Performativities and fabrications in the education economy: Towards the performative society. *Australian Educational Researcher, 17*(3), 1–24.

Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy, 18*(2), 215–228.

Ball, S. J. (2013). *Foucault, Power, and Education*. Routledge.

Banner, D. K. (1974). The politics of evaluation research. *OMEGA, The International Journal of Management Science, 2*(6), 763–774.

Beattie, L. (2020). Educational leadership: Producing docile bodies? A Foucauldian perspective on higher education. *Higher Education Quarterly, 74*(1), 98–110.

Berman, N. (2020). A critical examination of informal learning spaces. *Higher Education Research and Development, 39*(1), 127–140.

Botanic Gardens Conservation International (BGCI)/ Erasmus + (2019). Learn to engage - a modular course for botanic gardens. Erasmus + (Ref. 2016–1-UK01-KA202–024542). Retrieved 10th November 2022 from: https://erasmus-plus.ec.europa.eu/projects/search/details/2016-1-UK01-KA202-024542

Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352.

British Educational Research Association. (2018). *Ethical guidelines for educational research* (4th ed.). BERA.

Bulaitis, Z. H. (2020). *Impact and the humanities: The rise of accountability in public cultural life. Value and the humanities*. Palgrave Macmillan.

Camfield, L., & Duvendack, M. (2014). Impact evaluation-are we 'off the gold standard'? *The European Journal of Development Research, 26*(1), 1–11.

Collet-Sabé, J., & Ball, S. (2022). Beyond School. The Challenge of Co-Producing and Commoning a Different Episteme for Education. *Journal of Education Policy*. https://doi.org/10.1080/02680939.2022.2157890

Coultas, C. (2020). The performativity of monitoring and evaluation in international development interventions: Building a dialogical case study of evidence-making that situates 'the general.' *Culture and Psychology, 26*(1), 96–116.

Clapham, A. (2015). Producing the Docile Teacher: Analysing Local Area Under-Performance Inspection. *Cambridge Journal of Education, 45*(2), 265–280.
Clapham, A. (2016). Enacting informal science learning: Exploring the battle for informal learning. *British Journal of Educational Studies, 64*(4), 485–501.
Clapham, A., & Vickers, R. (2018). Further education sector governors as ethnographers: Five case studies. *Ethnography and Education, 13*(1), 34–51.
Cunningham, M., & Walton, G. (2016). *Informal learning spaces (ILS) in university libraries and their campuses: A Loughborough university case study*. New Library World.
Dawson, A., Yeomans, E., & Brown, E. R. (2018). Methodological challenges in education RCTs: Reflections from England's education endowment foundation. *Educational Research, 60*(3), 292–310.
de St Croix, T. (2020). *Re-imagining accountability: Storytelling workshops for evaluation in and beyond youth work* (pp. 1–18). Culture and Society.
Dondi, P., Porta, M., Donvito, A., & Volpe, G. (2022). A gaze-based interactive system to explore artwork imagery. *Journal on Multimodal User Interfaces, 16*(1), 55–67.
Dreyfus Hubert, L., & Rabinow, P. (1983). *Michel Foucault: Beyond structuralism and Hermeneutics*. Routledge.
Eckhard, S., & Jankauskas, V. (2020). Explaining the political use of evaluation in international organizations. *Policy Sciences, 53*(4), 667–695.
Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education, 26*(2), 247–273.
Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology, 16*(2), 171–190.
Foucault, M. (Ed.). (2000). *Essential works of Foucault: 1954–1984, Volume three, Power*. London: Penguin.
Falk, J. H., & Dierking, L. D. (2018). *Learning from museums*. Maryland: Rowman and Littlefield.
Falk, J. H. (2001). *Free-choice science education: How we learn science outside of school. Ways of knowing in science and mathematics series*. Teachers College Press.
Falk, J. H., & Dierking, L. D. (2010). School is not where most Americans learn most of their science. *American Scientist, 98*(6), 486.
Foucault, M. (1980). *Two lectures. Power/knowledge*. (ed/trans) C. Gordon. London: Longman.
Foucault, M. (1991). *Discipline and punish: The Birth of the Prison*. Penguin.
Foucault, M. (2010). *The government of self and others: Lectures at the Collège de France 1982–1983*. Palgrave Macmillan.
Garcia, B. (2012). What we do best. *Journal of Museum Education., 37*(2), 47–55.
Gerber, B. L., Marek, E. A., & Cavallo, A. M. L. (2001). Development of an informal learning opportunities assay. *International Journal of Science Education, 23*(6), 569–583.
Gewirtz, S., & Cribb, A. (2020). *Can teachers still be teachers? The near impossibility of humanity in the transactional workplace. Knowledge, policy and practice in education and the struggle for social justice: Essays inspired by the work of Geoff Whitty* (pp. 217). London: UCL Press.
Harrison, G. W. (2014). Impact evaluation and welfare evaluation. *The European Journal of Development Research, 26*(1), 39–45.
Howard, F. (2021). "It's Like Being Back in GCSE Art"—engaging with music, film-making and boardgames. Creative pedagogies within youth work education. *Education Sciences, 11*(8), 374.
Howard, F. (2022). *Global perspectives on youth arts programs: How and why the arts can make a difference*. Policy Press.
Jamieson, P. (2009). The serious matter of informal learning. *Planning for Higher Education, 37*(7), 18–25.
Jamieson, P. (2013). Reimagining space for learning in the university library. In G. Matthews & G. Walton (Eds.), *University libraries and space in the digital world* (pp. 142–154). Ashgate.
Jeffs, T., & Smith, M.K. (1999). Informal education. *Conversation, democracy and learning,* Ticknall: Education Now.
Jordan, H. (1994). Public parks, 1885–1914. *Garden History, 22*(1), 85–113.
Joslin, J. A. (2021). Capturing catalysis: A mixed-methods study raises questions on instrumentation's fit to mission. *Journal of Museum Education, 46*(3), 296–306.
Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Sage.
Lingard, B., Martino, W., & Rezai-Rashti, G. (2013). Testing regimes, accountabilities and education policy: Commensurate global and national developments. *Journal of Education Policy, 28*(5), 539–556.
Lyotard, J. (1984). *The postmodern condition: A report on knowledge*. Manchester University Press.

Maarschalk, J. (1988). Scientific literacy and informal science teaching. *Journal of Research in Science Teaching, 25*(2), 135–146.
Margolis, E., & Fram, S. (2007). Caught napping: images of surveillance, discipline and punishment on the body of the schoolchild. *History of Education, 36*(2), 191–211.
National Foundation for Educational Research NFER. (2022). Evaluation. Slough: NFER. Retrieved 13 January 2023, from: https://www.nfer.ac.uk/key-topics-expertise/research-methods-operations/evaluation/
Ozga, J. (2009). Governing education through data in England: From regulation to self-evaluation. *Journal of Education Policy, 24*(2), 149–162.
Ozga, J. (2020). The politics of accountability. *Journal of Educational Change, 21*(1), 19–35.
Pateman, C. (2012). Participatory democracy revisited. *Perspectives on Politics, 10*(1), 7–19.
Picciotto, R. (2014). Is impact evaluation evaluation? *The European Journal of Development Research, 26*(1), 31–38.
Pitts, P. (2018). Visitor to visitor learning: Setting up open-ended inquiry in an unstaffed space. *Journal of Museum Education, 43*(4), 306–315.
Quinn, J. (2018). Respecting young people's informal learning: Circumventing strategic policy evasions. *Policy Futures in Education, 16*(2), 144–155.
Rabinow, P., & Rose, N. (Eds.). (2003). *The essential Foucault: Selections from essential works of Foucault, 1954–1984*. The New Press.
Rende, K., Fromson, K., Jones, M. G., & Ennes, M. (2021). The privilege of low pay: Informal educators' perspectives on workforce equity and diversity. *Journal of Museum Education*, 46(4)
Rogoff, B., Callanan, M., Gutiérrez, K. D., & Erickson, F. (2016). The organization of informal learning. *Review of Research in Education, 40*(1), 356–401.
Sefton-Green, J. (2012). *Learning at not-school a review of study, theory, and advocacy for education in non-formal settings*. MIT Press.
Sim, N. (2019). *Youth work, galleries and the politics of partnership*. London. Palgrave McMillan
Sluga, H. (2011). 'Could you define the sense you give the word "political"'? Michel Foucault as a political philosopher. *History of the Human Sciences, 24*(4), 69–79.
Styles, B., & Torgerson, C. (2018). Randomised controlled trials (RCTs) in education research–methodological debates, questions, challenges. *Educational Research, 60*(3), 255–264.
Taylor, D., & Balloch, S. (2005). The politics of evaluation: An overview. *The politics of evaluation*, 1–18. Bristol; Bristol University Press.
Trainer, L. (2010). What is your museum's economic footprint? *Journal of Museum Education, 35*(3), 237–246.
van Voorst, S., & Mastenbroek, E. (2019). Evaluations as a decent knowledge base? Describing and explaining the quality of the European Commission's ex-post legislative evaluations. *Policy Sciences, 52*(4), 625–644.
Vestman, O., & Conner, R. (2006). *The relationship between evaluation and politics*. SAGE Publications Ltd
Weiss, C. H. (1993). Where politics and evaluation research meet. *Evaluation Practice, 14*(1), 93–106.
Wergin, J. F. (1976). The evaluation of organizational policy making. A political model. Review of Educational Research, 46(1), 75–115.