Unpredictable cocktails or recurring recipes? Identifying the patterns that shape collaborative performance summits

Scott Douglas & Marie-Jeanne Schiffelers

Published online: 25 Feb 2021.

Submit your article to this journal ⤤

Article views: 884

View related articles ⤤

View Crossmark data ⤤

Citing articles: 2 View citing articles ⤤

Routledge
Taylor & Francis Group

Unpredictable cocktails or recurring recipes? Identifying the patterns that shape collaborative performance summits

Scott Douglas and Marie-Jeanne Schiffelers

Utrecht School of Governance, Utrecht, The Netherlands

ABSTRACT
Advancing the performance of collaborations requires not only shared performance indicators, but also shared performance routines. Collaborative performance summits offer partners a routine for jointly explicating goals, exchanging performance information, examining progress, and exploring actions. However, summits can easily devolve into pointless talking shops or political warzones. Research has identified what ingredients shape a summit, but how exactly these ingredients interact and produce summit outcomes is less well understood. Through the systematic observation of eight summits, we identify and precisely describe 13 interaction patterns. These findings can be tested through future research and inform the design of summits.

Collaborative performance summits: Unpredictable cocktails?

Collaborative governance is popular among practitioners and scholars alike, but getting a grip on the performance of collaborations remains a challenge. Recent research efforts have made progress by identifying relevant performance *indicators* (e.g. Emerson and Nabatchi 2015; Page et al. 2015), but effective performance management also requires appropriate performance *routines* (Gerrish 2016; Moynihan and Kroll 2016). Performance routines such as shared goal-setting workshops, joint case discussions, and collective evaluations can help collaborative partners to explicate their goals, exchange performance information, examine the progress they are making, and explore potential actions for performance improvement (Douglas and Ansell, 2020).

A performance routine used frequently in collaborations is what Douglas and Ansell (2020) call a collaborative performance summit; the various partners in a collaboration gather in a room to collectively review their actions and results (cf. 'forums' in Bryson, Crosby, and Seo 2020, 'PerformanceStat meetings' in Behn 2014, 'learning forums' in Moynihan 2005). Similar to performance reviews inside organizations, collaborative performance summits serve as inter-organizational 'dialogue routines specifically focused on solution seeking, where actors collectively examine

---

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

information, consider its significance, and decide how it will affect future actions' (Moynihan, 2005; 33).

Collaborative performance summits can be instrumental in realizing better performance. Summits can help partners to make sense of performance data (Moynihan et al 2016), enable collaborative learning (Gerlak and Heikkila 2011), and motivate performance improvement (Moynihan and Kroll 2016). However, summits can also turn into pointless talking shops or give rise to conflict (Innes 1992). Although various studies have identified several key ingredients for fruitful summits, such as extensive preparation and able process guidance (Behn 2014), the exact dynamics of performance summits remain poorly understood (Moynihan, 2005; Laihonen and Mäntylä 2017).

Collaborative performance summits could be viewed as an attempt by participants to rationally solve performance puzzles, emphasizing the role of performance information 'as a stimulus for shared discussion, and [that a] performance dialogue is expected to result in an actionable solution' (Laihonen and Mäntylä 2017, 416). An opposite perspective would be that summits are fundamentally political arenas, where the political interests of the parties on the table will dominate the dynamics (Wildavsky, 1979; Bryson and Crosby 1993).

A more nuanced perspective would be that summits are formed by multiple interacting ingredients, with the mix of performance information, political interests, personal dynamics, etc. around the table shaping the summit dynamics. Moynihan (2006: 151–168) contends that these interactions are so complex and unpredictable that 'what one group of decision-makers concludes is a reasonable interpretation and an appropriate response may be completely at odds with another group's assessment. ... [D]ecision outcomes are unlikely to have a systematic relationship with performance information in a way that is easily observable to researchers'

We accept that the many and varied ingredients of a collaborative performance summit make for a complicated cocktail. However, we also contend that some recurring patterns can potentially be identified. Specific patterns may recur when observing multiple summits, where a specific configuration of performance information, process design, political interests, problem perspectives, and personal dynamics repeatedly gives rise to a specific outcome. These patterns are not likely to be simplistic and deterministic (more x equals more y), but multi-faceted and probabilistic (the presence of a in combination with the presence of b and c, increases the likelihood of d occurring) (Campbell 1966; Yin 2009).

We here aim to identify a first set of such recurring patterns by systematically observing eight collaborative performance summits. We aim to describe the precise mechanisms of each pattern so they can be tested as concrete propositions in future research. We start this endeavour with an overview of the key ingredients of collaborative summits as detailed by the literature. We then systemically observe the presence or absence of these ingredients in eight summits in the policy domain of adult illiteracy and trace how these configurations shape the summit dynamics. We identify 13 patterns which shape how summit participants explicate their goals, exchange information, examine information, and explore goals.

Dissecting collaborative performance summits

The ingredients of collaborative performance summits

The broader phenomenon of actors gathering to review their actions and results has been described from various perspectives, including deliberative democracy, collaborative policy-making, stakeholder consultation, and public management. Summits have subsequently been conceptualized as stakeholder conferences (Innes 1992), deliberative forums (Dryzek and Hendriks, 2012), goal review workshops (Bryson, Ackermann, and Eden 2016), PerformanceStat reviews (Behn 2014), or learning forums (Moynihan 2005; James et al. 2020).

Each of these different perspectives brings a different emphasis, our focus is to understand how summits can help improve the performance of a collaboration. The concept of collaborative performance summits as described by Douglas and Ansell (forthcoming) places summits at the nexus of collaborative governance and performance management. In this view, summits are part of the wider collaborative performance regime through which the partners determine and review their objectives.

Collaborative performance summits specifically are the 'interactive dialogue routines bringing together the various actors involved in a collaborative initiative to jointly (a) explicate their performance goals, (b) exchange performance information, (c) examine performance information, and (d) explore actions for potential performance improvement' (Douglas and Ansell, forthcoming: 7).

Across the various studies of performance summits, similar ingredients are identified as key to effective meetings. Most visibly, the *performance information* presented and examined at the summit influences its dynamics. Moynihan (2016) defines performance information as the combination of the objective (statistics, measurements), subjective (user experiences, employee experiences), and expert data (perspectives from scholars and trained professionals) brought to the discussion. Behn (2014) shows how network reviews benefit from well-prepared, well-structured performance information.

Equally visible is the role of *process design* in a summit; the sequence of the agenda items, the rules that govern the interactions, and the role of the chair or facilitator. Megnis and Eppler (2008), for example, show how the agenda design will influence who gets to speak and what is decided, while Quick and Sandfort (2014a) stress the role of the facilitator in creating safe and productive conversations.

A more subtle influence on the dynamics of a summit is the diversity of *perspectives on the problem* around the table (Laihonen and Mäntylä 2017). For example, doctors, teachers or parents participating in a summit discussing the fight against childhood obesity would all have a slightly different perspectives on the problem (Fishkin, 1991). On one hand, such differences may hinder a productive dialogue. On the other hand, the diversity in perspectives could help to identify new insights and opportunities for improvement (Ansell and Gash 2008).

The actors may also bring a diversity of *political interests* or organizational positions to the table (Moynihan 2005; Bryson and Crosby 1993). These various interests could hinder a frank discussion during the summit, but could also serve as an opportunity to build unexpected coalitions and reach multiple groups.

A final component, often hard to observe for external researchers, is the nature of the *personal relationships* between the individuals sitting around the table. Partners in a collaboration may have built up a history of distrust, conflict, and disparity over the

years, or face the challenge of building relationships with new partners (Gerlak and Heikkila 2011; Couturier and Sklavounos 2019).

A summit can lead to different outcomes, ranging from nothing much to big changes (Douglas and Ansell, forthcoming). In the first scenario, the energy built up during a summit dissipates after it is concluded and *no change* occurs (Innes 1992). A more productive summit may generate some *operational change*, with participants identifying opportunities for streamlining their daily actions (Gerlak and Heikkila 2011). More impactful summits lead to *strategic change*, where actors refocus the goals of the collaboration or change the way they monitor progress (Page et al. 2015).

Finally, summits can give rise to *constitutional change*, where the summits leads to drastic shifts in the makeup of collaboration as partners are removed or added and the fundamental decision-making rules governing the collaboration are redefined (Ostrom in McGinnis, 1989). Collaborative summits are shaped by the wider performance regime in which they take place, e.g. how partners usually work together will determine how they work together during the summit. Yet a summit may also shape the overall structure of the collaboration, i.e. decisions made during the summit could change how the collaboration is run. The overall collaborative structure shapes the summit practice, but the summit practice can shape the collaborative structure (Bryson, Crosby, and Seo 2020).

Figure 1 summarizes the different components of collaborative performance summits discussed. Building on the work of Douglas and Ansell (forthcoming), it places individual collaborative performance summits within the context of the wider collaborative performance regime, i.e. the collection of all the routines a collaboration uses to set and review its goals. The activities during the summit are shaped by the five ingredients discussed and can lead to four different outcomes.
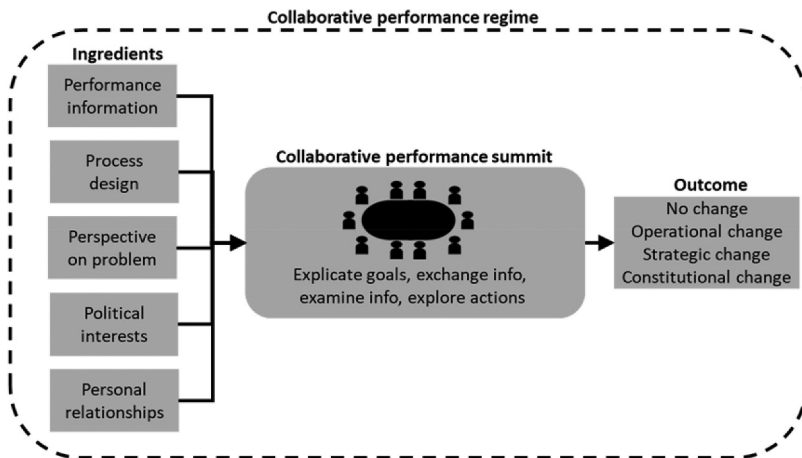

Figure 1. The ingredients, activities, and outcomes of collaborative performance summits.

The patterns of collaborative performance summits

There are different perspectives on how the various ingredients shape the activities and subsequent outcomes of a summit. In a straightforward view, a particular component can be regarded as dominant. For example, Wildavsky (1979) sees performance summits as political arena's where actors seek to maintain the status quo or advance their particular interest. In this view, performance information is hardly used and the political interests will be the primary shaping force.

By contrast, Behn (2014) frames summits should be primarily used as rational performance reviews to facilitate performance improvement. He argues that especially high-quality performance information and well-prepared process design will ensure successful meetings. In yet another perspective, Innes and Boohler (2010) argue that summits should be viewed as attempts at 'authentic conversations' where actors openly share their desires and concerns, making the personal relationships and process design the main drivers for effective summits.

More complex perspectives focus on the interplay between the various ingredients. Ansell and Gash (2008) view such summits as part of the collaborative process, which should be a virtuous cycle of building personal relationships, aligning political positions, jointly exploring performance information, and celebrating intermediate results. Moynihan (2006) argues that political interests and performance information interact as actors seek to bolster their arguments by citing the data that best supports their view. He concludes that this interaction is so complex that the outcome of a summit cannot be predicted beforehand, although it may be reconstructed afterwards.

We do accept that summits will involve complex interactions between the various ingredients, but also argue that some patterns can be expected to recur when observing multiple summits. For example, it would be reasonable to propose that collaborative summits where actors have strong, trust-based personal relationships would be more willing to exchange large amounts of performance information. Similarly, summits with more political antagonism between the participants would be less likely to feature an open exchange information and exploration of shared actions. However, the interaction between two ingredients of a summit – e.g. performance information and political interests – would probably also be affected by the other ingredients, such as the use stringent process design or presence of clashing perspectives on the problem.

The interaction between the various ingredients should therefore not be viewed as direct, linear relationship, but rather as nonlinear patterns where the different configurations of all the various ingredients together will generate specific summit dynamics and outcomes (Campbell 1966; Yin 2009). The patterns we expect to find are therefore multi-faceted and probabilistic (the presence of *abundant performance information* in combination with the presence of *limited political antagonism* and *shared perspectives on the problem*, increases the likelihood of *operational learning* occurring).

If we can formulate precise patterns we expect to occur, i.e. what exact configuration of ingredients we expect to produce a specific dynamic, we can test whether this pattern occurs in various summits. Using the systematic observation of eight summits, we generate a first set of patterns we would expect to occur in other summits as well. We do not aim to be *complete* in our description of summit patterns – that would require a large case collection encompassing all possible configurations of summit ingredients, activities, and outcomes – but we do aim to be *precise* in how we describe

the patterns we identify. Future research can then take these patterns and test them as propositions when observing other summits.

Methodology

Action research to generate theoretical propositions

The purpose of this research is to generate theoretical propositions from empirical observations about the highly complex phenomenon of summits. We needed to be able to get close to the data (i.e. observe the summits in action), observe comparable summits to see what difference changes in the ingredients make, and track the outcome of the summits over time as their impact may be delayed.

To satisfy these demands, we opted for an action research methodology, working with the Dutch Foundation for Adult Illiteracy to organize collaborative performance summits in eight municipalities across the country. In each municipality, we brought together a similar selection of partners (e.g. local government, community college, library, charities), took them through the same summit agenda, and collected the same data.

This action research methodology has the advantage of generating a lot of rich data, building theoretical propositions which are rooted in the complexity of real-life rather than the artificial constraints of a lab, and ensure that scientific insights generated connect with the experience of practitioners. However, action research is also problematic as it blurs the line between the observers and the observed and does not always generate traceable data (Brydon-Miller, Greenwood, and Maguire 2003).

We addressed some of the dangers by employing a mix of data collection methods; collecting data at different points in time and in different settings between researchers and subjects. However, the impact of the researchers on the subject cannot be discounted. The municipalities that agreed to invest time in a summit were probably already inclined to action, which was probably augmented by the pressure of having outside observers (see Hawthorne Experiment). For our purposes, the action research methodology provided a rich hunting ground for patterns, but the external validity of the patterns identified can only be verified through subsequent, more traditional research projects.

Case context

About 10% of the adult population in the Netherlands struggles with reading and writing, a rate comparable to most OECD countries. This functional illiteracy affects their ability to read letters from the government, follow medicine prescriptions, browse the internet, or manage their finances (National Illiteracy Foundation, 2019). Adult illiteracy is typical of a 'wicked issue' requiring collaboration (Head and Alford, 2015), as there can be many causes for the problem (migrant background, medical conditions, school dropout, etc.) and there is a need of a multiple, customized solutions (follow a formal course at community college, offer training through the employer, get a volunteer 'language buddy', etc.).

In the Netherlands, the local government is responsible for coordinating and financing most of the illiteracy programmes. Dutch local governments have an

extensive policy responsibility and relatively large administrative capacity. However, given the complexity of the issue, they still rely on the support of partners such as the local library, community college, job agency, and charities to find people in need of training and offer them the relevant support.

The adult illiteracy programs are financed by a grant for each municipality from the national government. The grants are determined by the population size and accompanied by restrictions on how the money can be spend, but there are no stringent performance objectives. Most municipalities pass this money to their partners, financing projects in the local library, paying for courses at the community college, subsidizing voluntary organizations to match people to language buddies, etc.

Most often, local governments define performance by the number of illiterate people reached through the various programs or by the number of people that successfully complete a formal language course. However, true to the wicked nature of the issue, this definition is often contested as charities argue that formal qualifications do not correspond to actual proficiency or match the needs of client groups (e.g. retired illiterate people often simply want to be able to read bedtime stories to their grandkids rather than complete a full course). In some municipalities, the local politicians are personally involved in setting and monitoring the performance targets for illiteracy programs, but often the illiteracy program get little structural attention and is left to civil servants.

Organizing eight summits

Working with the National Foundation for Illiteracy, a charity financed by the national government to promote illiteracy programs across the country, we organized collaborative performance summits in eight municipalities. We strove to get a relatively comparable set of municipalities; all municipalities are urban/semi-urban and have a sufficient population size to have a dedicated illiteracy program. However, local governments needed to volunteer themselves to host a summit. This probably means that those municipalities that participated were already motivated to improve their illiteracy collaborations.

Together with the local civil servant and a local representative of the National Foundation, we selected which partners should be invited to a particular summit. In each summit, the local government, library, and community college, and local project leader of the National Foundation participated. In some cases, these organizations choose to come with multiple representatives. Local charities (e.g. Salvation Army, volunteer associations) and social services featured in most summits as well. Job agencies were in attendance at half the summits and large employers joined two of them (see Table 1).

Every summit followed the same basic format: We as researchers opened and chaired meeting, restating that the purpose of the meeting was to jointly review the quality and performance of the local collaboration. The participants were then asked to select a picture from a range of photo's which best captured their view of the collaboration (i.e. a photo of a rowing team in sync versus a photo of a tug-of-war). Each participant was given a minute or two to explain why they choose that particular picture.

The participants were then presented with the anonymized results of a survey conducted among the participants prior to the summit, giving their assessment of

Table 1. Size of municipality in which the eight summits occurred and the summit participants.

#	Population	Summit participants							
		Local govt.	Comm. college	Library	National foundation	Charity	Job agency	Social services	Other
A	250.000	2	1	1	2		2		
B	270.000	1	2	1	1	2	1	1	2
C	150.000	3	2	2	1	2		1	
D	30.000	2	2	2	1			1	
E	60.000	1	1	1	1	2		1	
F	90.000	1	1	1	1	1	1		2
G	40.000	1	1	1	1		1	1	
H	60.000	1	1	1	1	1			

crucial aspects of the collaboration on a 1 to 5 scale (Are all relevant partners involved? Are the rules for collaboration clear? Is the collaboration achieving its goal?). Participants were asked to reflect on the outcomes of this survey to better understand how the collaboration was doing. Each meeting ended with an exploration of improvement actions. All summits lasted between one and a half and two and a half hours.

Observing the summits

We collected multiple types of data. Before the summit, each participant received a digital survey, asking them to rate aspects of the collaboration, which was discussed in anonymized version at the summit. We also conducted individual interviews with all of the participants of the first five summits to check whether we understood the concerns raised on the survey correctly. We took notes of the group discussion during each summit.

Six to twelve months after the event, we surveyed the local representative of the National Foundation to trace the impact of the summit on the collaboration. This representative was a relatively neutral observer of the local dynamics. As this check was done some time after the summit and other forces may have shaped the collaboration in the meantime, any changes can never be attributed purely to the summits.

We operationalized the ingredients, activities, and outcomes of the collaborative summits to systematically compare what occurred at each summit. For each of the ingredients and activities, we identified the possible configurations (e.g. there could be limited, moderate, or extensive performance information available, limited, moderate, or extensive exploration of future actions during the summit). This operationalization of complex phenomena will always be a simplification – e.g. the quality of the personal dynamics is reduced to whether actors did or did not know each other previously – but did provide us with a basic framework to structure our observations (see Tables 2, 3 and 4).

Based on the different sources of data we coded the ingredients, activities, and outcomes of each summit, where the two researchers involved would check each other's classification. For example, in the survey for summit H participants indicated that there was extensive *performance information* available about the collaboration,

Table 2. Operationalization of ingredients.

Ingredients		Limited	Moderate	Extensive
Performance information	Performance information available to all participants participating in the summit	• Limited info available	● Moderate info available	● Extensive info available
Process design	Active process guidance from facilitator during summit	• Limited guidance	● Moderate guidance	● Extensive guidance
Perspective on problem	Convergence of participating partners in their views on the societal challenge	• Limited convergence	● Moderate convergence	● Extensive convergence
Political positions	Convergence of participating partners in their interests and incentives	• Limited convergence	● Moderate convergence	● Extensive convergence
Personal dynamics	Familiarity among participating partners from knowing each other and seeing each other frequently	• Limited familiarity	● Moderate familiarity	● Extensive familiarity

Table 3. Operationalization of activities.

Activities		Limited	Moderate	Extensive
Explicating goals	Discussion on the content and nature of the goals for the collaboration	• Limited explication	● Moderate explication	● Extensive explication
Exchanging information	Swap of performance information (data, experiences, case examples) between participants	• Limited exchange	● Moderate exchange	● Extensive exchange
Examining information	Discussion on assessing the implications of performance information	• Limited examination	● Moderate examination	● Extensive examination
Exploring actions	Discussion on exploring actions to improve	• Limited exploration	● Moderate exploration	● Extensive exploration

Table 4. Operationalization of outcomes.

Outcomes		No change	Operational	Strategic	Constitutional
Outcome of summit	Report from local National Foundation representative on changes/continuities in local illiteracy collaboration	Collaboration has made no changes to activities or ambitions	Collaboration changed daily activities and services	Collaboration changed performance goals and monitoring	Collaboration changed participants and governance of collaboration

signalling they were well aware of each other's actions and capabilities, but did feel that information about the overall program results were missing.

The interviews before summit H showed that the participants had a convergent *problem perspective*, all arguing for a cross-sector, customized approach. The interviews showed moderately divergent *political interests*, with especially the library and community college clashing over the best allocation of funds, but all partners agreeing that a joint approach would ultimately benefit all parties.

The *personal dynamics* between the actors in summit H were characterized by high-trust, as actors did report in the survey to work together closely over a long period of time and felt able to raise difficult topics (as indeed happened during the summit). The role of the facilitator *process design* of the summit itself was active, but mainly in making sure the rapid discussion kept on track and all necessary topics were covered in the time available.

During summit H, there was an extensive *explication of the goals* and *exchange of information*, which was interesting as the goals of this collaboration were formalized in several pacts and the partners had previously indicated already knowing a lot about each other. All participants at the table supported the conclusion that even though the network had been doing well in hitting its initial goals and had a good view of its current activities (reaching migrants who lacked Dutch skills and get them to formal education), it had to advance to targeting more complex problems (reaching native Dutch speakers who feel ashamed about their illiteracy and get them customized help).

The wide-ranging discussion and broad information exchange enabled a relatively quick and shared *examination of the information* (a consensus that they were under-performing in reaching native speakers) and *exploration of future actions* (investing in a systematized approach on this front). Twelve months later, the local representative of the National Foundation reported that this summit had been part of a *strategic change* in the collaboration, integrating extra goals in the illiteracy program and mobilizing the social services and job agency in the illiteracy drive to better target native speakers.

We tracked similar data for all eight of the summits. Table 5 presents all of the cases, where the size of the dot indicates the extent to which an ingredient was present (availability of performance information, active process design, convergence in problem perspective, convergence in political interests, familiarity in personal dynamics) and an activity occurred during the summit (explication of goals, exchange of information, examination of information, exploration of actions).

Table 5. Overview ingredients, activities, and outcomes for each summit.

#	Ingredients					Activities				Outcomes
	Perform. Info	Process design	Perspect. problem	Political interests	Personal dynamic	Explicate goals	Exchange info	Examine info	Explore actions	
A	•	•	•	•	•	•	•	●	●	Constitutional change, shift to multilateral steering
B	•	●	•	●	•	•	•	●	⬤	Constitutional change, shift to local govt. taking lead
C	•	⬤	•	●	●	•	⬤	•	●	Constitutional change, local govt taking active role
D	•	●	•	●	●	•	●	⬤	⬤	No change yet, intentions to improve strategy
E	•	●	●	●	●	•	●	●	⬤	No change yet, intentions to improve strategy
F	●	•	●	●	●	●	●	●	⬤	No change, intentions to improve operations
G	●	⬤	⬤	●	⬤	⬤	⬤	●	●	Strategic change, with new priorities for collaboration
H	⬤	⬤	⬤	●	⬤	⬤	⬤	●	●	Strategic change, with new priorities for collaboration

•= limited ● = moderate ⬤ = extensive presence

The cases are listed based on the extent to which the summit ingredients and activities occurred. Cases which had little ingredients present and very limited discussions come first (e.g. case A and B), while summits which had both extensive supplies of performance information, convergence of interests, etc. and spent extensive time exchanging information, exploring actions, etc. are listed towards the end (e.g. case G and H).

Identifying patterns

Explicating goals: Enforced harmony, superficial agreement, or deepening discussion

We first observed how extensively the summit participants would discuss the goals of the collaboration and to what extent they converged or diverged in this process of goal explication. At first sight, there were two types of patterns; summits where everyone agreed quickly on the goals and summits where goals were hotly contested. A closer analysis showed there were more subtle dynamics at play.

At the summit in town A, little time was spent on the goals. 'Well, we're here to reduce adult illiteracy, with a particular focus on targeting native Dutch speakers' said the senior official of the local government, thereby ignoring other potential goals for the network. In the individual interviews before the summit, the local partners had stated that they actually knew very little about the purpose of the collaboration overall, but as the lead official (who was also in charge of allocating the money) pushed her point, the other participants did not speak up to question her assessment.

The summit in town B also quickly skipped past the goals discussion. Here a few dominant actors said it was unnecessary to discuss the goals at length as it should be obvious. Again, other participants had said in the individuals interviews beforehand and indicated in the survey that the purpose was unclear. As one education partner said: 'There is no shared ambition, no shared direction, and therefore I have no idea what results we are aiming for'. Yet they did not voice their doubts in the summit setting.

Even though many participants lacked performance information and their perspectives on the problem where strongly divergent, the hierarchical political positions and low-trust personal dynamics at these summits meant that the dominant actors would lay down their interpretation of the goals, with everyone nodding along in enforced harmony.

During the summits in the towns of C, D, and E, the explication of the goals was equally short, but for different reasons. The pre-summit survey had shown that the actors had very little ideas about the purpose of the network, just as they were unclear about the governance structures, each other's activities, and the results of the collaboration. There was a silence in the room at summit C when participants were asked to share their view on the goals, until one participant said: 'I guess we are all working on adult illiteracy, sort of'. The other participants nodded along. Divergent perspectives on the problem and limited performance information here enabled only superficial agreement on the goals.

By contrast, at summit H, and to a lesser extent summits F and G, the question about the goals of the collaboration immediately prompted a spirited debate about the purpose of the collaboration. A charity director in H kicked off the discussion with:

'We focus too much on reading skills, we should focus more on the quality of life of these participants.' A participant in summit F broadened the scope of the collaboration by stating 'Language lessons alone are not the solution for poverty and social exclusion. We need to do more to address these bigger issues.'

Interestingly, the interviews and surveys in F, G, and H had indicated that these participants were relatively more aligned in their perspectives on the problem and were actually more clear and explicit about the goals of the collaboration than the actors in the other collaborations. In other words, the relative convergence of problem perspective, abundance of performance information, resulted in richer debates about the goals of the collaboration, enabled by the close personal dynamics (see Table 6).

Table 6. Three patterns observed in the explication of goals during summits.

Patterns while explicating goals during summits
Enforced harmony
Insufficient *performance information* + diverging and hierarchical *political positions* + low trust *personal relationships* Participants reaffirm the goals of collaboration as laid out by dominant actors
Cases: A, to some extent B
Superficial agreement
Insufficient *performance information* + relatively divergent *problem perspectives* + moderately trusting *personal relationships* = Participants are only able to formulate a superficial agreement on goals
Cases: C, D, E
Deepening debate
Abundant *performance information* + moderately trusting *personal relationships* + moderately converging *perspectives on the problem* = Participants explore and challenge goals of the collaboration
Cases: F, G, H

Exchanging information: Hoarding, prospecting, or mining information

At summits A and B, the discussion moved briskly from the brief exploration of goals to a quick exchange of the available performance information about the collaboration. Even though the majority of participants had indicated in the survey they lacked a good overview of the collaboration and its activities, little extra information was shared during the summit. One or two actors would dominate the discussion, asking people to share specific items of information (e.g. how do you target illiteracy among the unemployed?) rather than inviting a broad information exchange (e.g. what does your organization do in the domain of illiteracy?)

Even interventions by the process facilitators, asking people to share more information, generated little information as participants seemed hesitant to share data. Moreover, the dominant actors would repeatedly censor information (e.g. 'That is not relevant right now.' or 'Those are just your impressions.'). Despite a widely reported lack of performance information, the antagonistic political positions and low-trust personal dynamics drove people to hoard the information they had.

The summits C, D and E had spent little time explicating their goals, achieving only superficial agreement on some vaguely stated goals, but did dedicate a large share of their meeting to exchanging performance information. This information was mainly focused on getting a grip on the basics, with participants swapping operational facts: Who is in charge of the policy? What are the different organizations doing? Who allocates the budget?

This exchange was generally valued by the participants as key lacunas and misconceptions were addressed. However, the discussion did not begin to cover information beyond the basics. What methods were people trying to reach illiterate adults? What results did these methods generate? What potential for synergy was there between the various activities of partners? The participants felt safe enough to share their basic information, but were still prospecting in the dark for what information might be relevant.

The summits in F, G, and H, spent a large proportion of the meeting time on exchanging information. Especially for H this could be considered surprising, as the respondents had indicated in the survey and interviews that they did already have quite a good overview, yet even in this network people were asking each other for still more detail and background information. Having extensive information seems to encourage an appetite for yet more information.

Compared to summit C, D, and E the information exchange in these three collaborations was more purpose-driven and fast-paced. Participants would focus on a particular issue (e.g. the link between literacy and health in G, the link between literacy and social exclusion in H) and then quickly mined all the information available from among the participants of the summit. (see Table 7).

Table 7. Patterns observed in the exchange of information during summits.

Patterns while exchanging information during summits
Hoarding information in unsafe environments
Low trust *personal relationships* + strongly antagonistic *political positions* = Participants withhold *performance information*
Cases: A, B
Prospecting information in unknown territory
Little existing *performance information* + at least moderate trust-based *personal relationships* = Participants will share basic information but struggle to determine what is relevant
Cases: C, D, E
Mining information in safe environments and known territory
Available *performance information* + high trust *personal relationships* + at least moderately unified *political positions* = Participants will extensively exchange yet more information about specific themes
Cases: F, G, H

Examining information: Marching, jumping, or moving slowly towards conclusions

The summits then turned to examining the state of the collaboration. Summit A and B spent a fair share of their meeting time drawing conclusions about the collaboration, even though the participants had indicated in the survey that they actually had little information to base such judgements on and little information was exchanged during the summit itself.

Rather than lining up the relevant information and crafting the relevant assessments, the dominant partners would steer the discussion by being the first to formulate their assessment, with the other participants rarely voicing dissent or alternative opinions. For example, the survey in summit A reported that all executive partners felt there was insufficient money to their job effectively, but the leading actor said 'I think we can all agree that money is not the problem here.' and closed the discussion there.

In summit D, the group was also eager to formulate assessments about the collaboration, even though their shared information position was limited and their goals were only superficially explicated. These summits seemed prone to jumping to conclusions based on little information. For example, the participants would propose new modes for finding and training illiterate adults even though there was no evidence that the current methods were not working or that the new methods would be any better.

The participants of the summits in B, C, E, F, and G were more hesitant to formulate assessments. 'This is too early to tell, we should collect more information.' was the most repeated phrase during this stage in the summits. They were slow to move towards conclusions out of necessity, as they lacked the key information.

In summit H, where actors actually had the most information available, partners were hesitant to come to strongly worded assessments. Participants would caution each other that experience had warned them to be careful to come to sweeping conclusions. 'This is interesting, let's see if we can get more information and let's make concrete plans then.' The partners in H were modest not out of necessity but out of experience (see Table 8).

Table 8. Patterns observed in the examining of information during summits.

Patterns while examining information during summits
Marching to conclusions
Antagonistic *political positions* + low trust *personal relationships* = Dominant actors force the summit towards judgements not necessarily supported by the performance information
Cases: A, to some extent B
Jumping to conclusions
Little *performance information* + moderate convergence in *political interests* = Actors will come to quick assessments not necessarily supported by the performance information
Cases: A, to some extent B
Stumbling slowly towards conclusions
Little *performance information* + moderately diverging *perspectives on the problem* + moderately trusting *personal relationships* = Participants feel unable to formulate any assessments
Case: C, E, F
Plotting deliberately towards conclusions
Abundance of *performance information* + moderately converging *perspectives on the problem* + strong *personal* relationships = Participants will still be cautious to formulate strong assessment as they make plans to get more information
Case: H, to some extent G

Exploring actions: Overpowering, overpuzzling, small but significant steps

Each summit concluded with exploring potential follow-up actions. All sessions ended with the participants saying that this discussion had been useful and the group should meet again, even if the summit had actually been quite tense and arguably not very informative. Participants acted like family members saying at the end of a tense family Christmas dinner 'let's make sure it wont be a year until we see each other again' and then being relieved it is over. Looking more closely, again revealed more subtle differences in the final phase of the different summits.

In case A and B, the exploration of future actions was again dominated by a select set of actors dictating what actions to take next. In B, the small group of dominant actors outlined the marching orders for the collaboration as a whole, only allowing for some minor tweaks to the strategic goals of the collaboration. However, as the representative of the National Foundation later reported, the local government concluded after the meeting that the divide between the small group of actors and the rest

of the partners could not continue. The local government decided to change the constitution of the collaboration, by taking a more active leadership role and working to involve all partners equally.

At summit A, the lead official who had dominated proceedings throughout came to a similar conclusion during the summit itself. Although previously a supporter of bilateral performance reviews between the local government and each separate partner, this lead actor at the end of the summit proposed to stop having bilateral meetings as they were divisive. Instead, the official proposed that the illiteracy initiative should be steered more multilaterally. This shift was enacted by scheduling (and indeed organizing) a yearly governance meetings with all partners represented.

At summit C, D, E, and F, where the proceedings were hampered by a scarcity of information though the personal relationships were relatively relaxed, the final phase at first generated little concrete next steps. At summits D and E, the discussion had been so wide-ranging, covering so many basic points, that even active process intervention could not pinpoint potential next steps. The local representatives reported that no changes were made yet in these networks, even though some intentions to make operational improvements were still being debated.

At summit C, where the exchange of information had explicated a clear need for more leadership from the local government, the discussion of the next steps stuck to more general points about 'communicating more with each other'. Here a process intervention from the facilitators ('But what are you going to do about the lack of leadership?') prompted the civil servants on the table to commit to taking a lead role, which was indeed confirmed to be happening by the local foundation representative. A similar dynamic occurred at summit F, where there were some clear operational bottlenecks identified during the exchange of information, but the closing discussion avoided this point until the facilitator brought them back into focus. However, no real changes were observed afterwards.

In summit G and H, which were the meetings with the best-informed participants and most convergence on goals and political positions, participants were very most modest in deciding next steps. Even though the discussion had been wide-ranging and participants were committed to change, the agreed-to actions were modest ('let's meet and discuss a draft plan'). However, the follow-up observation revealed that these networks actually had made concrete steps in the months after the summits, agreeing on new strategic visions (case G) or even implementing them already (case H). They took small but significant and deliberate steps towards performance improvement (see Table 9).

Table 9. Patterns observed in the exploring of actions during summits.

Patterns while exploring actions during summits
Overpowering
Antagonistic *political positions* + low trust *personal relationships* = Dominant actors will force the summit towards specific future actions, although this could also be towards more multilateral steering
Cases: A and B
Overpuzzling
Little shared *performance information* + moderately divergent *perspectives on the* problem + no dominants actors in leading *political positions* = Participants are prone to analysis paralysis
Cases: C, D, E, F
Small but significant steps
Extensive *performance information* + converging *perspectives on the problem* + moderately converging *political positions* = Participants move towards slow but significant changes
Case: G and H

Conclusion and discussion

Summits are an often used intervention for jointly reviewing the performance of collaborations, but to maximize their potential and minimize their downsides, we must understand what makes summits work. This systemized study of the ingredients and dynamics of eight collaborative performance summits generated a first set of propositions about how summits work. This list of propositions is but a first overview based on a relatively small set of cases in a particular policy context and is intended to be tested and refined through future research.

The list of 13 patterns supports a more nuanced perspective on the dynamics of collaborative performance summits, advocated already by the likes of Moynihan (2006) and Ansell and Gash (2008). Summits are not purely political (Wildavsky, 1979) or rational activities (Behn 2014), but complex social events where ingredients such as performance information, process design, problem perspective, political positions, and personal relationships interact and jointly produce the dynamics at the meeting and subsequent outcomes.

Small changes can make a big difference in this cocktail, as nuanced shifts in the configuration of ingredients can make a big difference in the pathway a summit will follow. The participants in both summit A and C had little performance information available, but where in A this was combined with antagonistic political positions and the summit devolved into tense and limited exchange, in summit C this lack of performance information was combined with more trusting relationships and the summit became an open exchange of what knowledge partners did have.

Even though the interaction may be multi-faceted and complex, the sharpness of the propositions suggests that the ingredients and dynamics of a summit are not wholly unknowable. The presence or absence of the ingredients occurs in specific configurations that shape the activities in the summit and how the summits goes on to influence the collaboration overall (Bryson, Crosby, and Seo 2020). Each summit maybe a complex cocktail of ingredients, but some recurring recipes can potentially be identified.

Methodological challenges

These tentative ideas must be considered in the light of the limitations of the research design of this study in particular and the methodological challenges of studying summits in general. In this study, the patterns identified could be the product of the specific policy field (i.e. illiteracy), level of government (i.e. local), population size (smaller localities do better than big localities), or country setting (i.e. the Netherlands). Summits in other contexts may produce different patterns, requiring replication research across different settings (Walker, James, and Brewer 2017).

The researchers themselves were a prominent ingredient in this summit in their role as process facilitators (Brydon-Miller, Greenwood, and Maguire 2003). Moreover, there was a large decree of coder discretion in how the researchers labelled the presence or absence of ingredients, even though two researchers reviewed each other's codings. Replicating this method through other researchers and through more regimented data collection methods could help to further the first insights collected here.

Furthermore, although key elements of the cases were kept consistent (format of summit, level of government, etc.), there was variation in the cases (size of

municipality, makeup of participants, precise professional and organizational background of participants). The impact of changing these ingredients will have to be examined in future more extensive case comparison designs.

This study also raised a more fundamental methodological challenge for the study of summits. Often, the patterns observed looked very similar at first glance, revealing crucial differences only when studied in more detail. For example, both summit B and D spent very little time on explicating the goals, but for very different reasons. In summit B, there was no safe environment to explore differences in priorities, where in D the actors were simply lost in unknown territory. This difference only became observable when looking at the summit dynamics in full.

A counterintuitive observation was also that collaborations with the least information (e.g. C) were more keen to come to sweeping conclusions than collaborations with the most information (e.g. G). This suggests a collective version of the Dunning Kruger effect where people with low competence display more confidence than people with high competence (Kruger and Dunning 1999), again requiring a keen and critical eye of the observing researcher.

Such subtle differences in patterns only become apparent when studying the summits in full and tracing their impact over time. This potentially makes one-off, experimental studies seeking to isolate one ingredient of summits difficult, demanding the collection of a lot of background information from multiple sources to understand just one summit.

Future research

Moving forward, there are multiple strands of research required to further unpack the dynamics of collaborative performance summits. This first exercise generated 13 propositions about patterns that occur at summits. Further exploratory studies in different contexts could seek to complete this list of propositions, striving to reach a saturation point as any additional studies do not generate additional patterns.

A second strand of research could seek to confirm whether these patterns do indeed occur in other summits. Single case studies could use strict pattern matching (Yin 2009) to establish whether the patterns occurred as expected in a single summit. Medium N studies could apply Qualitative Case Analysis or statistical tests to explore the causal relationships between the different ingredients, activities, and outcomes. Large-scale data exercises and controlled experimental settings could be used to corroborate or reject the propositions forwarded.

A final strand of research could explore what interventions can help to bolster or break patterns, exploring what changes in the performance information, process design and other summit ingredients could positively influence the dynamics of the summit. Across these efforts, we may not expect that summits will ever become fully predictable, replicable events, but can strive to increase our insight in what recipes make for a good cocktail.

Acknowledgments

Disclosure statement

Funding

Notes on contributors

Scott Douglas, DPhil is an assistant professor at the Utrecht School of Governance. He specializes in the performance management of collaborations dealing with complex issues such as radicalization, illiteracy, and obesity.

Dr. Marie-Jeanne Schiffelers is an assistant professor at the Utrecht School of Governance. Her work centres on supporting governments in dealing with complex societal challenges in the domains of security, public health, and education.

References

Ansell, C., and A. Gash. 2008. "Collaborative Governance in Theory and Practice." *Journal Of Public Administration Research And Theory* 18 (4): 543–571. doi:10.1093/jopart/mum032.

Behn, R. D. 2014. *The PerformanceStat Potential: A Leadership Strategy for Producing Results.* Washington DC: Brookings Institution Press.

Bititci, U., P. Garengo, V. Dörfler, and S. Nudurupati. 2012. "Performance Measurement: Challenges for Tomorrow." *International Journal of Management Reviews* 14 (3): 305–327. doi:10.1111/j.1468-2370.2011.00318.x.

Brydon-Miller, M., D. Greenwood, and P. Maguire. 2003. "'Why Action Research?'." *Action Research* 2003/1 (p): 9–28. doi:10.1177/14767503030011002.

Bryson, J. M., and B. C. Crosby. 1993. "Policy Planning and the Design and Use of Forums, Arenas, and Courts." *Environment and Planning. B, Planning & Design* 20 (2): 175–194. doi:10.1068/b200175.

Bryson, J. M., B. C. Crosby, and D. Seo. 2020. "Using a Design Approach to Create Collaborative Governance." *Policy and Politics* 48 (1): 167–189.

Bryson, J. M., F. Ackermann, and C. Eden. 2016. "Discovering Collaborative Advantage: The Contributions of Goal Categories and Visual Strategy Mapping." *Public Administration Review* 76 (6): 912–925.

Campbell, D. T. 1966. "Pattern Matching as an Essential in Distal Knowing." In *The Psychology of Egon Brunsvik*, edited by K. R. Hammond, 81–106. New York: Holt, Rinehart & Winston.

Couturier, J., and N. Sklavounos. 2019. "Performance Dialogue: A Framework to Enhance the Effectiveness of Performance Measurement Systems." *International Journal of Productivity and Performance Management* 68 (4): 699–720. doi:10.1108/IJPPM-09-2017-0238.

Douglas, S., and C. Ansell. 2020. "Getting a grip on the performance of collaborations: Examining collaborative performance regimes and collaborative performance summits." Public Administration Review. https://doi.org/10.1111/puar.13341

Dryzek, J., and C. Hendriks. 2012. Fostering deliberation in the forum and beyond. In *The argumentative turn revisited: Public policy as communicative practice*, 31–57. Duke University Press.

Emerson, K., and T. Nabatchi. 2015. "Evaluating the Productivity of Collaborative Governance Regimes: A Performance Matrix." *Public Performance and Management Review* 38 (4): 717–747. doi:10.1080/15309576.2015.1031016.

Fishkin, James S. 1991. *Democracy and deliberation: New directions for democratic reform (Vol. 217).* New Haven, CT: Yale University Press

Gerlak, A. K., and T. Heikkila. 2011. "Building a Theory of Learning in Collaboratives: Evidence from the Everglades Restoration Program." *Journal of Public Administration Research and Theory* 21 (4): 619–644. doi:10.1093/jopart/muq089.

Gerrish, E. 2016. "The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis." *Public Administration Review* 76 (1): 48–66. doi:10.1111/puar.12433.

Head, Brian W., and John Alford. 2015. "Wicked problems: Implications for public policy and management." *Administration & society* 47 (6): 711–739.

Innes, J. E. 1992. "Group Processes and the Social Construction of Growth Management: Florida, Vermont, and New Jersey." *Journal of the American Planning Association* 58 (4): 440–453. doi:10.1080/01944369208975828.

Innes, Judith E., and David E. Booher. 2010. *Planning with complexity: An introduction to collaborative rationality for public policy*. Routledge.

James, O., A. L. Olsen, D. Moynihan, and G. G. Van Ryzin. 2020. *Behavioral Public Performance: How People Make Sense of Government Metrics*. Cambridge: Cambridge University Press.

Kruger, J., and D. Dunning. 1999. "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-assessments." *Journal of Personality and Social Psychology* 77 (6): 1121. doi:10.1037/0022-3514.77.6.1121.

Laihonen, H., and S. Mäntylä. 2017. "Principles of Performance Dialogue in Public Administration." *International Journal of Public Sector Management* 30 (5): 414–428. doi:10.1108/IJPSM-09-2016-0149.

Mengis, J., and M. J. Eppler. 2008. "Understanding and managing conversations from a knowledge perspective: an analysis of the roles and rules of face-to-face conversations in organizations." Organization Studies 29 (10): 1287–1313.

Moynihan, D. P. 2005. "Goal-based Learning and the Future of Performance Management." *Public Administration Review* 65 (2): 203–216. doi:10.1111/j.1540-6210.2005.00445.x.

Moynihan, D. P. 2006. "What Do We Talk about When We Talk about Performance? Dialogue Theory and Performance Budgeting." *Journal of Public Administration Research and Theory* 16 (2): 151–168. doi:10.1093/jopart/muj003.

Moynihan, D. P., and A. Kroll. 2016. "Performance Management Routines that Work? an Early Assessment of the GPRA Modernization Act." *Public Administration Review* 76 (2): 314–323. doi:10.1111/puar.12434.

O'Toole, J. L. J., and K. J. Meier. 2004. "Public Management in Intergovernmental Networks: Matching Structural Networks and Managerial Networking." *Journal Of Public Administration Research And Theory* 14 (4): 469–494. doi:10.1093/jopart/muh032.

Ostrom, E. 1989. "Microconstitutional change in multiconstitutional political systems." *Rationality and Society* 1 (1): 11–50.

Page, S. B., M. M. Stone, J. M. Bryson, and B. C. Crosby. 2015. "Public Value Creation by Cross-Sector Collaborations: A Framework and Challenges of Assessment." *Public Administration* 93 (3): 715–732. doi:10.1111/padm.12161.

Quick, K., and J. Sandfort. 2014a. "Learning to Facilitate Deliberation: Practicing the Art of Hosting." *Critical Policy Studies* 8 (3): 300–322. doi:10.1080/19460171.2014.912959.

Quick, K. S., and J. Sandfort. 2014b. "Learning to Facilitate Deliberation: Practicing the Art of Hosting." *Critical Policy Studies 8, No* 3 (3): 300–322

Walker, R. M., O. James, and G. A. Brewer. 2017. "Replication, Experiments and Knowledge in Public Management Research." *Public Management Review* 19 (9): 1221–1234. doi:10.1080/14719037.2017.1282003.

Wildavsky, A. 1979. "Policy as its own cause". In *The Art and Craft of Policy Analysis*, 62–85. London: Palgrave Macmillan.

Yin, R. K. 2009. "How to Do Better Case Studies." *The SAGE Handbook of Applied Social Research Methods* 2: 254–282.