
Part 3: **Assessment**

Didier Caluwaerts and Min Reuchamps

18 Evaluating citizens' assemblies: Criteria, methods and tools

Abstract: Citizens' assemblies (CAs) have been heralded as a potential antidote to the alleged crisis of democratic legitimacy and effectiveness. However, CAs are used in a variety of designs, on a variety of topics, and in a variety of political and institutional contexts. As the experience with CAs deepens and competing designs proliferate, demands for concrete evaluation standards and tools have risen. This chapter aims to offer a comprehensive account of central evaluation criteria for CAs, both in the input, throughput, and output phases, and it gives an overview of potential methods for evaluating CAs. Moreover, the chapter also outlines a new operational evaluation tool, the CA Evaluation Survey (CAES), which taps into the different dimensions.

Keywords: citizens' assemblies, deliberative democracy, legitimacy, evaluation, institutional designs, methods, input, throughput, output, CA Evaluation Survey (CAES)

18.1 Introduction

Citizens' assemblies (CAs) have witnessed a significant rise in the last decade as a potential antidote to the alleged crisis of democratic legitimacy and effectiveness. Despite its common denominator, however, CAs are used in a variety of designs, on a variety of topics, and in a variety of political and institutional contexts. As the experience with CAs deepens and competing designs proliferate, demands for concrete evaluation standards and tools have risen. After all, the more power is given to CAs, the more we expect them to live up to certain quality standards.

In the wake of the *Catching the Deliberative Wave* report (2020), the OECD has proposed *Evaluation Guidelines for Representative Deliberative Processes* (2021). The framework relies on a twofold evaluation: process evaluation and outcome evaluation. The former taps on two dimensions: process design integrity ("Evaluating the design process that set up the deliberation") and deliberative experience ("Evaluating how a deliberative process unfolds 'in the room' and 'outside the room'"). The latter looks at the pathways to policy impact ("Evaluating influential conclusions and/or actions of a deliberative process") and CAs transformative effects on the wider public ("the secondary and long-term effects on efficacy and public attitudes").

This chapter builds on the OECD's criteria and on previous frameworks for evaluating CAs. Its aim is to offer a comprehensive account of central evaluation criteria for CAs, in the input, throughput, and output phases (Caluwaerts and Reuchamps

Didier Caluwaerts: Vrije Universiteit Brussel, Belgium; **Min Reuchamps:** Université catholique de Louvain, Belgium.

2015; 2016), and it gives an overview of potential methods for evaluating CAs. Moreover, the chapter also outlines a new operational evaluation tool, the *Citizens' Assembly Evaluation Survey* (CAES),¹ which taps into the different dimensions. In the following sections, we will systematically deal with the input, throughput, and output dimensions of CAs, and offer some core questions evaluators should ask when assessing the quality of CAs. A full version of the CAES can be found in the appendix.

18.2 Input

The first dimension to assess a CA relates to input. Three main criteria can be distinguished. The first one taps into the representativeness of the CA, that is to what extent the composition of the CA reflects the diversity of the society in which the CA takes place. The openness of the agenda comes as a second criterion: it is the question of what is to be discussed by the CA and who decides it. The third criterion is epistemic completion, that is to what extent the participants have access to information and the quality of this information.

18.2.1 Representativeness

CAs aim to gather a public that is ideationally and discursively representative of the wider societal diversity. It is often argued that a key dimension of deliberative democracy is the nature of the *representativeness* it allows. This echoes the principle of “all-affected interests” (Dahl 1970: 49–63), which means that everyone affected by the issue at stake should be included in the deliberation (Young 2000). It entails a twofold question: who are the participants, and how are they recruited?

Several answers can be given to the question of who the participants are and thus how a CA accommodates representativeness. Two main ways of conceiving representativeness are usually put forward. On the one hand, it could call for socio-demographic representativeness, that is trying to reach a descriptive representation of the society (Griffin et al. 2015). On the other hand, it could also call for discursive or epistemic representativeness that would insist on a diversity of opinions and discourses rather than on a diversity of people themselves (Landemore 2013). In assessing the representativeness of a CA, it could be a mixture of both.

In order to assess representativeness, participant surveys are the most often used tools with a focus on socio-demographics in order to compare the sample of participants to the population at large (for the socio-demographic representativeness) and/

¹ Even though the CAES has been developed for the evaluation of citizens' assemblies and other mini-publics, most of its criteria can be applied to other (more aggregative) types of participatory governance as well.

or on issue-based items (for the discursive representativeness) (Caluwaerts and Reu-champs 2015). Discourse analysis can also be mobilized to assess the discursive diversity among participants during the deliberation. This technique allows to dig further but it is more time and energy consuming. It also requires either a direct observation of the deliberations and/or recording them for a post-hoc analysis. Additionally, the evaluators can focus on citizen perceptions of the diversity of the group or on the question whether they feel like they were in contact with diverse viewpoint. Possible CAES items include:

- “The opinions of the other participants on [the issue under discussion] did not fundamentally differ from my own.”
- “As a consequence of my participation in the CA, I have come to understand different viewpoints on [the issue under discussion].”

The technique used to delimit the *who* should also be under close scrutiny, as its consequences for representativeness can vary a lot depending on whether random selection, targeted selection, or self-selection is in operation (Caluwaerts and Ugarriza 2012; Fung 2006). CAs seek representativeness. It is therefore key to assess not only whether it is achieved but also how it is achieved. The recruitment can indeed be – partly – biased, in particular among traditionally vulnerable groups.

In order to assess the recruitment, an analysis of the script of the selection procedure and recruitment efforts can be performed. In the script, the recruitment procedure should be detailed so that any observer can understand the choices that have been made and possibly their consequences. Some follow-up individual interviews with participants and non-participants could also shed light on this criterion. Non-participants (who are citizens who have been invited but have either declined the invitation or have accepted it but are not part of the final sample) need to be included in the assessment as they provide a view on the recruitment process less biased by their own participation (Devillers et al. 2020).

18.2.2 Openness of the agenda

Another key criterion of the input dimension is *agenda-setting*, that is, the *what* question. Fung has distinguished between “hot” and “cold” issues: the former might be of greater social and political interest but more productive of tension; the latter may provoke less tension but have less social resonance, within both the CA and society at large (Fung 2007). Above all, how the agenda is decided on is of crucial importance to understanding the dynamics of any CA. Several ways of agenda-setting, each fostering different consequences, can be found in practice. The process could have an open agenda: the entire population or all stakeholders are able/invited to set or vote on the agenda in an open-ended process. At the other end of the spectrum, the agenda could be closed, decided on by formal institutions and with little room for introducing new issues. In-between options are also available. For instance, the agenda could be fixed but partic-

ipants could also be allowed to introduce adjacent issues, and question whether pre-chosen issues should be on the agenda at all. As we will see later on, the choice for an open or closed agenda comes with inevitable trade-offs. The more open and inclusive the agenda-setting process, the harder it will prove to be to generate political up-take.

In order to assess the openness of the agenda, we suggest two tools. On the one hand, a document analysis can be performed in order to apprehend the scope of the mandate given to the CA. This is an external assessment of this criterion. On the other hand, an internal assessment of this criterion is needed, that is how participants perceive the openness of the agenda. In this perspective, the CAES contains some items related to the agenda and ask the participants to position themselves on this question. These include:

- “I feel like I had a genuine say in which issues could be discussed.”
- “Even though the CA dealt with [the issue under discussion], I felt we could discuss adjacent issues.”

18.2.3 Epistemic completeness

A last dimension of the input dimension is the level of information of the people who deliberate, which can be referred to as “epistemic completeness” (Mucciaroni and Quirk 2006). More specifically, it is not so much what participants know but how they can learn about the issues at stake. Ideally, participants have access to all relevant information and are – made – competent, with access to experts on the question, policymakers and/or resource people. In practice, however, efforts to inform citizens may be limited to information booklets, with little room for extra learning and questioning (Fishkin & Luskin 2005). We can also see here the interactions – and possibly the trade-offs – between the different criteria of the input dimension. For instance, if the aim is to build a CA involving a large and diverse crowd of people, the question of information is even more important, to ensure sufficient epistemic completeness. Above all, while there are interactions within the input dimension, there are also interactions *between* input, throughput, and output dimensions, to which we will turn in the remainder of this chapter:

In order to assess the epistemic completeness, we also suggest an external approach and an internal approach. A media analysis can be used to assess the epistemic completeness from an external point of view. Media reports offer a comprehensive assessment of the information given to participants but also the potential biases. Next to this external approach, participants themselves should be consulted about what they think of the quality of the information they have received. The CAES contains some items related to epistemic completeness as well as the role and influence from the experts met throughout the CA. These include, among others:

- “The information materials/the experts presented balanced views on the [the issue under discussion].”

- “The information materials/the experts helped me to better understand the challenges involved in [the issue under discussion].”

18.3 Throughput

A second important step concerns evaluating the process of deliberation and decision-making itself. This so-called throughput phase is concerned with the distribution of power within the group and the external power exerted over the participants. More specifically, we focus on three complementary criteria for evaluating the process. These are the quality of participation, the quality of decision-making, and the contextual independence of the process.

18.3.1 Quality of participation

As we highlighted above, high quality CAs are diverse and inclusive. They include a group that is by-and-large representative of the wider population or of those affected by the deliberation. However, an assembly with a descriptively representative participant pool is not necessarily substantively representative of the wider maxi-public. It can only be so when the voices that are present are also allowed to be voiced, and this is where participatory quality and equality comes in. A CA can only truly be called deliberative when it is designed in such a way as to bring out everyone's arguments experience and perspectives, and to foster openness towards the arguments of others. In this respect, the quality of participation criterion captures whether the discussion reflects the characteristics of the ideal speech situation (Bekkers and Edwards 2007), even though deliberative democrats have repeatedly argued for a wider definition of the type of discourse that should be considered deliberative. Bächtiger et al. (2010) argue for instance that deliberation should be open to personal experiences, humour, storytelling and rhetoric (see also Dryzek 2000).

Crucial in this respect is participatory equality. Even though broad inclusion is normatively ideal, it could also bring with it some practical problems. After all, some might not be able to live up to the high cognitive and intellectual demands of good deliberation, and others who master the vernacular of good argumentation might use this to dominate the deliberations and impose their own views, simply because they are formulated in a more eloquent manner (Lupia and Norton 2017).

Discursive inequalities can thus undermine the quality and legitimacy of CAs, and one of the most important tasks for deliberative designers consists of massaging out these inequalities. Several simple group-dynamic techniques can help realize the goal of participatory equality. These include, among others, varying group sizes (from large plenary groups to small one-on-one deliberations because large groups have exclusionary tendencies), varying moderator roles (from neutral guarantors of the deliberative rules to more active devil's advocates), and allowing sufficient time

for deliberation (more time creates more opportunities to speak) (Caluwaerts and Reuchamps 2015). Other characteristics of CAs (e.g., multilingualism or the level of polarization on the issue under discussion) can also increase inequalities.

Assessing the quality of participation can happen in several ways. First of all, most – if not all – CAs are based on a detailed script outlining the time slots, the goals, and the group dynamic aspects. A script analysis can give the evaluator some first clues as to the group dynamic qualities of the deliberations. However, a genuine evaluation of the participatory quality and equality will inevitably require non-participant observation. After all, formal scripts and formally delineated moderator roles do not tell us much about how deliberation materializes in practice. Equality on paper more often than not results in inequality in practice. Finally, the CAES might be a particularly useful tool for mapping participants' perceptions of the deliberative qualities of the discussions. If participants felt like they were silenced or ridiculed, the CAES survey will inevitably pick up on dysfunctional group dynamics and poor participatory quality. This can be measured using some of the following items:

- “During the discussion, I had ample opportunity to express my views.”
- “The moderators did not manage to eliminate all inequalities between the participants.”
- “We were offered sufficient time to discuss [the issue under discussion].”

18.3.2 Quality of decision-making

CAs are based on the principle that deliberation can lead to epistemically better and democratically more legitimate decisions than voting. It is therefore important that decisions in a deliberative setting come about through a process of argumentation. Decisions made through CAs must reflect the reasoned opinion and openness to persuasion of all those involved and not the power relations in the group (Caluwaerts and Reuchamps 2016).

However, the aim of reaching some kind of uncoercive consensus, which is central to deliberative theory (Dryzek 2000; Steiner 2012; Steiner et al. 2004), is at odds with the realities of political decision-making. When dealing with salient political issues and when much is at stake, conflicts within a mini-public will inevitably continue to linger. The theoretical ideal of the naturally occurring consensus hits some practical obstacles when deciding on actual policies. If this is the case, the organizers will have to resort to other techniques, and – despite the conflicting logics of aggregation and deliberation – a balance should be struck between talking and voting in the real world of deliberative CAs (Saward 2000: 67–68).

Most deliberative designs therefore rely on some aggregative mechanisms to come to some kind of final decision, so that the power of the majority does play a role. This should not necessarily undermine the quality of decision-making as long as the phase of deliberation outweighs the voting phases, and as long as there is an opportunity for the minority to make its opinions and arguments heard in the final recommendations.

The quality of decision-making thus depends on the sheer number of times aggregation crosses the deliberative process and the binding power of these aggregative decisions. For instance, regularly interrupting the deliberative process for a straw poll to see whether and how opinions have shifted, gives the initial majority group more influence on the process than a final vote at the end of the event. Starting off the CA with a vote might lock people into their initial positions since they want to avoid losing face by switching sides. In contrast, allowing those participants on the losing end of a vote to present their dissenting opinion in the final recommendations, increases the quality of decision-making. Similarly, the type of rule might impact the quality of decision-making as well. After all, super-majority rules have previously been found to lead to more high-quality deliberation than simple majorities, and might therefore increase the legitimacy of the decision (Caluwaerts and Deschouwer 2014).

The assessment of the quality of decision-making in essence boils down to an assessment of the relative weight of deliberation and aggregation in CAs. Evaluators should determine how often, with which rules, at which times, and with what effect the assembly relies on voting procedures. This can – similarly to the quality of participation – be done based on script analysis. The design scripts should give a clear overview of organizers' reliance on voting. Additionally, evaluators should look at the final report with the recommendations and policy proposals. This can offer some indication of whether the dissenting opinions within the participant group are included. Finally, the CAES could be useful in assessing participants' perceptions of ownership during the process and in the final outputs, with items such as:

- “I feel like the final recommendations reflect all of the participants' ideas.”
- “My own ideas about [the issue under discussion] are sufficiently reflected in the final recommendations.”

18.3.3 Contextual independence

Finally, evaluators should not only assess the internal power dynamics within the assembly; external coercion is undesirable as well for the legitimacy of CAs (Caluwaerts and Reuchamps 2015). Deliberation is more legitimate if no outside pressure is exerted on the participants, but at the same time, CAs and their participants are under severe public scrutiny. They often deal with controversial issues lacking any type of political consensus, and their participants are a relatively small subset of a wider population that is not necessarily well-acquainted with deliberative modes of political decision-making. This embeddedness in a wider political ecosystem could create tensions and outside pressures (Caluwaerts and Reuchamps 2016). A legitimate deliberative process should therefore be able to handle these outside influences and should avoid the participants from being forced in a particular course of action. Deliberants should thus be substantively independent from these outside pressures exerted by political parties, public opinion makers, pressure groups or the media.

Evaluating the independence of the CA from external pressures is no mean feat because pressure can be exerted in many different ways. One possible source of data would consist of a (social) media analysis. Media reporting and social media posts in the run up or during the CA are oftentimes revealing because they indicate which actors (politicians, civil society groups...) want the recommendations of the participants to go in which direction. Additionally, the participants might have experienced pressure outside of the public and visible channels. A thorough evaluation would therefore also require an assessment of citizens' perceptions via the CAES:

- “I feel that I was pressured into agreeing with recommendations I do not fully endorse.”
- “I did not present my own views on [the issue under discussion] because I thought it was expected by politicians or the media.”

18.4 Output

With the output dimension, we come to the key fact that a CA does not work in isolation but takes place in a broader public sphere and political system. It is therefore crucial to assess how a CA fits into this larger spectrum. That is how the society at large takes up the issues raised by the CA. In assessing this dimension, three criteria can be distinguished. First, we suggest assessing public endorsement for the CA and its recommendations. Second, the political uptake needs also to be looked at. Third, the assessment of the political uptake should also be complemented by an assessment of the policy implementation or lack thereof.

18.4.1 Public endorsement

As Dryzek insightfully wrote over twenty years ago: “decisions still have to be justified to those who did not participate” (Dryzek 2001: 654). Indeed, how the public perceives the CA and its recommendations is an important facet of any assessment of a given CA (see the chapter by Rountree and Curato in this *Handbook*). In fact, a CA is a mini-public of the larger so-called maxi-public, or put it simply the public. Even if one does not go as far as Fishkin's argument of “what the public *would* think, if it had a more adequate chance to think about the questions at issue” (Fishkin 1992: 26, emphasis in the original), some form of public endorsement is expected for a CA. Research on CAs on electoral reform in Canada and in the Netherlands has shown that large portions of the society were not aware of the deliberative democracy processes in motion (Fournier et al. 2011).

In order to assess public endorsement, the oft-used tool is a general population survey asking a representative sample whether they have heard about the CA, what they think of it, and how much they support its recommendations. To deepen the understanding of the public endorsement or the lack thereof, respondents' opinions vis-à-

vis the CA could be analysed in light of their socio-demographics and political and democratic attitudes as well as policy preferences on the questions at issue. While this direct measurement of public endorsement is needed, an indirect approach can be helpful. A script analysis can highlight the efforts of the organizers and the means they have set to maximize public endorsement, and the following CAES items can measure the participants' perceptions:

- “I shared the recommendations of the CA via mail or social media.”
- “I am convinced that a large part of the citizens can endorse the recommendations of the CA.”

18.4.2 Political uptake

Next to public endorsement, political uptake is also key to assess (see also the chapter by Minsart and Jacquet in this *Handbook*). By political uptake, we mean the discussion of the CA and its results in conventional political arenas: parliament, government, political parties (Jacquet et al. 2016). Here a distinction needs to be made between political actors at large and, if relevant, those who are the initiators of the CA. For the former, it is useful to assess their awareness and views of the CA, as it was done for the public. For the latter, one could gauge the evolution of their support through time, but it is also particularly relevant to understand their reasons to set a CA in motion and how they envision the follow-up. In fact, some forms of feedback can be foreseen; report on the decisions and progress made to participants and the general public, or only on demand by participants.

In order to assess political uptake, a survey among political actors can be run. When relevant, it can take the same form as the general population survey. It can be complemented by discourse analysis broadly understood. Political actors produce several forms of discourse: in parliament, in the media, in manifestos, to give three main examples. Elite interviews offer another approach to assess political uptake as it always digs into rationales, values, and visions of the CA by the political actors. Finally, the perceptions of the participants might be polled via the following CAES items:

- “It was clear from the start what was going to happen with the recommendations of the CA.”
- “I am convinced that politicians will take into account the CAs' recommendations.”

18.4.3 Policy implementation

Policy implementation can be seen as the final step of a CA, but not all CAs reach this stage, as the chapter by Minsart and Jacquet argues in this *Handbook*. Most CAs yield recommendations but in most cases there is no automatic implementation. It can be related to the fact that it was not politically foreseen by its initiators or because these recommendations cannot be translated as they are into policies. This criterion

focuses on the potential policy implementation brought about by the CA. Such assessment needs to take into account two difficulties: scope and time (Caluwaerts and Reuchamps 2015). On the one hand, it is seldom the case that recommendations are translated into one single policy and rather the implementation is achieved via several policies. On the other hand, implementation takes time, and therefore the assessment needs to take this variable into account.

In order to assess policy implementation, the first suggestion is to perform a policy documentation analysis, seeking to trace the fate of each recommendation. Because the policy implementation processes are multifaceted and include several actors at different stages, Vrydagh and Caluwaerts (2020) propose the Sequential Impact Matrix that distinguishes between the type of influence (continuous, enriching, shifting, innovating, and inhibiting) and the extent of influence (no uptake, partial uptake, full uptake). A systematic and comprehensive approach is needed to fully grasp the policy implementation criterion in assessing any CA, and the citizens' perceptions thereof can be measured with the following items:

- “I am convinced that we will find my ideas in tomorrow's policies.”
- “I am convinced that we will find the CAs' recommendations in tomorrow's policies.”

18.5 Discussion

The criteria outlined above give a comprehensive and theoretically grounded assessment of the overall qualities of a CA. However, the lingering question among most evaluators of deliberative processes is whether all criteria should be met or not in order for CAs to be legitimate. For instance, would we prefer a CA with high levels of representativeness and impact and low levels of deliberation, over a highly deliberative process among likeminded individuals and with low impact? These are normative trade-offs that every evaluator will inevitably encounter.

Several practical considerations are therefore in order. First of all, when conflicting results occur, we do not advocate treating every criterion with the same weight. Normatively, we consider the quality of representation (input), the quality of participation (throughput), and the political uptake (output) to be the primary evaluation criteria. Gross disregard of one of these criteria should be considered a fundamental flaw in the deliberative design of CAs. Secondly, evaluators should refrain from applying these criteria in a binary fashion (either a criterion is respected or disregarded). The evaluation of CAs requires a more fine-grained approach that determines whether some criteria are met to some degree, and that also takes into account contextual factors, such as the overall financial support or the saliency of the topic. Finally, some reflections on positionality are in order. Evaluators should be reflexive of their own role and position within the organization of the CA, of their own socio-demographic and ideational characteristics, and the impact of these characteristics on their observations. They should also avoid one-sided assessments by ideally combining a multitude of data

and perspectives. High quality evaluation ideally builds on the triple foundations of non-participant observation, document analysis, and the application of the CAES.

A final consideration is that criteria are also (practically and politically) inter-linked. Design choices implemented to meet one criterion, might affect the evaluation of other criteria as well (Caluwaerts and Reuchamps 2016). For instance, the output dimension works in close interaction with the composition of the CA and of agenda-setting on the input side. Choices made to increase representativeness or the openness of the agenda have an impact on the likely political uptake. All in all, the expert evaluators of CAs should bear in mind that trade-offs are inherent in democracy and democratic innovations, and that any evaluation should be explicit about which instruments are used and how they are calibrated.

References

- Bächtiger, A., Niemeyer, S., Neblo, M., Steenbergen, M. R., & Steiner, J. (2010). Disentangling diversity in deliberative democracy: Competing theories, their blind spots and complementarities. *Journal of Political Philosophy* 18 (1), 32–63.
- Bekkers, V., & Edwards, A. (2007). Legitimacy and democracy: A conceptual framework for assessing governance practices. In V. Bekkers, G. Dijkstra, A. Edwards, & M. Fenger (eds), *Governance and the Democratic Deficit. Assessing the Legitimacy of Governance Practices*, 35–60. Aldershot: Ashgate.
- Caluwaerts, D., & Deschouwer, K. (2014). Building bridges across political divides: Experiments on deliberative democracy in deeply divided Belgium. *European Political Science Review* 6 (3), 427–450.
- Caluwaerts, D., & Ugarriza, J. E. (2012). Favorable conditions to epistemic validity in deliberative experiments: A methodological assessment. *Journal of Public Deliberation* 8 (1).
- Caluwaerts, D., & Reuchamps, M. (2015). Strengthening democracy through bottom-up deliberation: An assessment of the internal legitimacy of the G1000 project. *Acta Politica* 50 (2), 151–170.
- Caluwaerts, D., & Reuchamps, M. (2016). Generating democratic legitimacy through deliberative innovations: The role of embeddedness and disruptiveness. *Representation* 52 (1), 13–27.
- Dahl, R. A. (1970). *After the Revolution? Authority in a Good Society*. New Haven: Yale University Press.
- Devillers, S., Vrydagh, J., Caluwaerts, D., & Reuchamps, M. (2020). Invited but not selected: The perceptions of a mini-public by randomly invited – but not selected – citizens. *ConstDelib Working Paper Series* (4), 1–21. Retrieved from <https://constdelib.com/wp-content/uploads/2020/02/WP4-2020-v.2-CA17135.pdf>
- Dryzek, J. S. (2000). *Deliberative Democracy and Beyond. Liberals, Critics, Contestations*. Oxford: Oxford University Press.
- Dryzek, J. S. (2001). Legitimacy and economy in deliberative democracy. *Political Theory* 29 (5), 651–669.
- Fishkin, J. S. (1992). The idea of a deliberative opinion poll. *Public Perspective* 3 (2), 26–27.
- Fishkin, J. S., & Luskin, R. C. (2005). Experimenting with a democratic ideal: Deliberative polling and public opinion. *Acta Politica* 40, 284–298.
- Fournier, P., van der Kolk, H., Carty, R. K., Blais, A., & Rose, J. (2011). *When Citizens Decide: Lessons from Citizens' Assemblies on Electoral Reform*. Oxford: Oxford University Press.
- Fung, A. (2006). Varieties of participation in complex governance. *Public Administration Review* 66 (S1), 66–75.
- Fung, A. (2007). Minipublics: Deliberative designs and their consequences. In S. W. Rosenberg (ed.), *Deliberation, Participation and Democracy: Can the People Govern?*, 159–183. Basingstoke and New York: Palgrave Macmillan.

- Griffin, J., Abdel-Monem, T., Tomkins, A., Richardson, A., & Jorgensen, S. (2015). Understanding participant representativeness in deliberative events: A case study comparing probability and non-probability recruitment strategies. *Journal of Public Deliberation* 11(1). doi: <https://doi.org/10.16997/jdd.221>
- Jacquet, V., Moskovic, J., Caluwaerts, D., & Reuchamps, M. (2016). The macro political uptake of the G1000 in Belgium. In M. Reuchamps & J. Suiter (eds), *Constitutional Deliberative Democracy in Europe*, 53–73. Colchester: ECPR Press.
- Landemore, H. (2013). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton, NJ: Princeton University Press.
- Lupia, A., & Norton, A. (2017). Inequality is always in the room: Language & power in deliberative democracy. *Daedalus* 146 (3), 64–76.
- Mucciaroni, G., & Quirk, P. J. (2006). *Deliberative Choices: Debating Public Policy in Congress*. Chicago: The University of Chicago Press.
- Saward, M. (2000). Less than meets the eye: Democratic legitimacy and deliberative theory. In M. Saward (ed.), *Democratic Innovation: Deliberation, Representation and Association*, 66–77. London: ECPR Studies in European Political Science.
- Steiner, J. (2012). *The Foundations of Deliberative Democracy: Empirical Research and Normative Implications*. Cambridge: Cambridge University Press.
- Steiner, J., Bächtiger, A., Spörndli, M., & Steenbergen, M. R. (2004). *Deliberative Politics in Action. Analysing Parliamentary Discourse*. Cambridge: Cambridge University Press.
- Vrydagh, J., & Caluwaerts, D. (2020). How do mini-publics affect public policy? Disentangling the influences of a mini-public on public policy using the Sequential Impact Matrix Framework. *Representation* 1–20. doi:10.1080/00344893.2020.1862901
- Young, I. M. (2000). *Inclusion and Democracy*. Oxford: Oxford University Press.

Appendix: An evaluative tool for CAs: The Citizens' Assembly Evaluation Survey (CAES)

Dimension	Criterion	CAES items (Likert scale answering categories)
Input	Representativeness	– The participants in the CA were broadly representative of the wider population. – A broad range of perspectives on [the issue under discussion] was present in the participant group. – The opinions of the other participants on [the issue under discussion] did not fundamentally differ from my own. – As a consequence of my participation in the CA, I have come to understand different viewpoints on [the issue under discussion].
	Openness of the agenda	– I feel like I had a genuine say in which issues could be discussed. – Even though the CA dealt with [the issue under discussion], I felt we could discuss adjacent issues. – The organizers did not allow us to broaden the discussion to other relevant issues. – The moderators did not allow us to broaden the discussion to other relevant issues.
	Epistemic completeness	– The information materials presented balanced views on the issue under discussion. – The experts offered balanced views on the issue under discussion. – The information materials presented diverse arguments for and against [the issue under discussion]. – The experts presented diverse arguments for and against [the issue under discussion]. – The experts during the CA were well chosen. – Did you look up information yourself on [the issue under discussion] before or during your participation in the CA? – As a consequence of my participation in the CA, my knowledge on [the issue under discussion] has improved. – As a consequence of my participation in the CA, I have come to understand other perspectives on [the issue under discussion]. – The experts helped me to better understand the challenges involved in [the issue under discussion]. – The experts used clear and understandable language. – I learned a lot about [the issue under discussion] from my participation in the CA. – I had sufficient information about [the issue under discussion] to engage in a meaningful discussion with the other participants.
Throughput	Quality of participation	– During the discussion, I had ample opportunity to express my views. – No one dominated the discussion and everyone had an opportunity to speak.

Continued

Dimension	Criterion	CAES items (Likert scale answering categories)
		- During the discussion, many people just stated positions without justifying them. (inversely coded) - I feel that I was needlessly interrupted during the discussion. (inversely coded) - No matter how hard I tried, the other participants seemed unwilling to listen to what I had to say. (inversely coded) - Most participants seemed to care only about their own well-being. (inversely coded) - I think most of the participants in the discussion tried hard to make a compelling case for their viewpoints. - We did not have sufficient time to discuss all issues. - I was able to present my opinion on [the issue under discussion] in a nuanced manner within the timeframe of the CA. - The moderators did not manage to eliminate all inequalities between the participants. - Some topics that I thought were important were not covered during the discussions. - We were offered sufficient time to discuss [the issue under discussion]. - We were offered sufficient time to interview the experts. - We were offered sufficient time to write down the final recommendations. - It was clear from the start what was expected of me. - It was clear from the start what will happen with the recommendations we formulated.
	Quality of decision-making	- I feel like the final recommendations reflect all of the participants' ideas. - Even though there was a vote in the end, I do feel that our group discussions are reflected in the final recommendations. - My own ideas about [the issue under discussion] are sufficiently reflected in the final recommendations.
	Contextual independence	- I feel that I was pressured into agreeing with recommendations I do not fully endorse. - I did not present my own views on [the issue under discussion] because I thought it was expected by politicians. - I did not present my own views on [the issue under discussion] because I thought it was expected by the media.
Output	Public endorsement	- I discussed [the issue under discussion] with others outside of the CA (e.g. friends, family, colleagues). - I shared the recommendations of the CA via mail or social media. - I am convinced that a large part of the citizens can endorse the recommendations of the CA. - I am convinced that a large part of the politicians can endorse the recommendations of the CA. - I am convinced that a large part of the policy makers can endorse the recommendations of the CA.

Continued

Dimension	Criterion	CAES items (Likert scale answering categories)
	Political uptake	– It was clear from the start what was going to happen with the recommendations of the CA. – I am convinced that politicians will take into account the CA's recommendations.
	Policy implementation	– I am convinced that we will find my ideas in tomorrow's policies. – I am convinced that we will find the CA's recommendations in tomorrow's policies.

