Methods for Analyzing and Measuring Group Deliberation

Laura W. Black
Ohio University

Stephanie Burkhalter
Humboldt State University

John Gastil
University of Washington

Jennifer Stromer-Galley
University of Albany, SUNY

Paper developed for inclusion in L. Holbert (ed.),
Sourcebook of Political Communication Research:
Methods, Measures, and Analytical Techniques
(New York: Routledge).
Methods for Analyzing and Measuring Group Deliberation

Abstract

This chapter describes the range of methods scholars use to measure and analyze deliberation in small groups. Although deliberative democracy extends beyond the small group, our overview is limited to group discussion, which is the foundation for deliberative theory. Contemporary theorists conceptualize deliberation as a communication process in which groups engage in a rigorous analysis of the issues at hand and also engage in a social process that emphasizes equality and mutual respect. Much of the scholarship in this area uses indirect measures of deliberation by focusing on antecedents or outcomes such as attitude change, depth of understanding, and argument repertoire. Attempts to directly measure aspects of deliberation include asking participants for assessments of their experiences, closely analyzing the content of ostensibly deliberative meetings, and/or conducting holistic case studies of deliberative groups. This chapter discusses the strengths and weaknesses of each of these approaches and offers methodological recommendations for the future study of group deliberation.
Supporters and skeptics of deliberation have recently called for increased empirical research of the process as it is used in small groups (Delli Carpini, Lomax, & Jacobs, 2004; Fung, 2007; Gastil, 2008; Levine, Fung, & Gastil, 2005; Rosenberg, 2005, 2006). There exists a pressing need for studies of actual deliberation informed by social science methodology because numerous civic minded organizations have already put into place small group deliberations guided by the premises of deliberative theory (see Gastil & Levine, 2005 for examples; Mansbridge, et. al. 2006).\(^1\) A challenge for any scholar attempting to measure public deliberation is to translate normative concepts such as analytic rigor, equality, respect, and consideration into measurable variables that are flexible enough to recognize deliberation in a variety of practical settings. Despite wide ranging enthusiasm for the process and a proliferation of deliberative projects, there exists no universally accepted—or even conventionally adopted—definition of small group public deliberation, let alone a consistent set of empirical measures derived from such a definition (see Burkhalter, Gastil, & Kelshaw, 2002; Delli Carpini et al., 2004; Stromer-Galley, 2007).

In the absence of a consistent definition of small group public deliberation, researchers have drawn on diverse literatures to create research designs aimed at empirically capturing aspects of deliberation. In this chapter, we discuss examples of research that operationalizes normative conceptions of democratic deliberative theory to gauge the deliberativeness of small group discussion.\(^2\)

\(^1\) Since the focus of these practical deliberations tends to be on a particular decision-making task or outcome, the organizations that run them do not regularly provide empirical studies of the process itself. Self-study of discussion process is also not part of their funding mandate, nor do staff necessarily have program evaluation training or traditions. The Deliberative Democracy Consortium (www.deliberative-democracy.net) aims to rectify this by bringing together researchers and practitioners.

\(^2\) In this chapter, we do not aim to review all of the empirical studies of deliberation completed to date. Good reviews include Delli Carpini et al. (2004), Mendelberg (2002), and Ryfe (2005). The online *Journal of Public Deliberation* is a good resource for examples of post-2004 empirical studies of deliberation.
We begin with a brief explanation of a model of small group deliberation proposed by Burkhalter et al. (2002) and elaborated on by Gastil and Black (2008) that has provided a basis for measures in some of these studies. In our subsequent discussion of measures of deliberation, we divide them into two main types: those that directly measure aspects of a theoretical definition of deliberation (direct measures) and those that attempt to measure deliberation by studying variables that can be seen as indicators of deliberative processes (indirect measures). In the conclusion, we argue that a more rapid advance in deliberation research requires a more consistent definition and corresponding operationalizations of small group deliberation.

Empirical research of small group deliberation remains in its early stages, and we expect that future studies will refine considerably the measures that we review herein. In this chapter, we aim simply to make the wider community of deliberation scholars aware of the full range of approaches to measurement currently in use.

A model of small group public deliberation

Burkhalter et al. (2002) provide a conceptual definition and theoretical model of public deliberation in small groups that we think is a good starting point for the development of reliable measures that could be widely used to expand the body of empirical research on deliberation.\(^3\) Their model defines ideal small group deliberation as “a combination of careful problem analysis and an egalitarian process in which participants have adequate speaking opportunities and engage in attentive listening or dialogue that bridges divergent ways of speaking and knowing” (p. 398). An innovation of this model is that group discussion is seen as more or less deliberative depending on how deliberative norms are put into practice and how participants respond to those norms. The basic norms of deliberation are analytic rigor concerning the problem and potential

---

\(^3\) This is perhaps not surprising since two of the authors of this chapter collaborated to develop the model!
solutions under discussion, careful and respectful consideration of information and diverse points of view on the issue, provision of sufficient opportunities for participants to speak, and recognition of—though not necessarily agreement with—participants’ different approaches to speaking and understanding.

Burkhalter, et. al.’s (2002) model identifies deliberative norms as potentially measurable discussion conditions that, if met, could produce certain outcomes. These outcomes include mainly post-deliberation changes among the participants, such as reinforcing deliberative habits, the discovery of previously unrecognized shared values and identities, an increased sense of democratic citizenship (or community identification), increased analytic and communicative skills necessary for political reasoning, and increased feelings of political efficacy. The authors also theorize that contextual characteristics, such as motivation to deliberate and the perceived potential for common ground, will influence the group’s ability to deliberate well.

In a recent work, Gastil and Black (2008) build on Burkhalter et al.’s (2002) conceptualization to give a more detailed explanation of the analytic and social processes involved in public deliberation. As they describe it, deliberation’s analytic processes include building an information base, prioritizing key values, identifying solutions, weighing solutions, and making the best decision possible (if the situation calls for a decision). The social processes include ensuring equal and adequate speaking opportunities to group members, demonstrating mutual comprehension of one another’s perspectives, adequately considering the views of other participants, and demonstrating respect for one another.4

The definitions advanced in Burkhalter et al. (2002) and Gastil and Black (2008) provide concepts that can be operationalized in empirical studies to examine deliberative processes and

4 Gastil and Black (2008) and Gastil (2008) demonstrate how this definition can be extended meaningfully to characterize deliberative practices beyond the small group setting, from dyadic conversations to governments, media systems, electoral processes, and societies.
Measuring Group Deliberation - 5

consequences. However, even when researchers create high-quality measures, additional challenges to the study of group deliberation remain. For instance, because theorists conventionally understand deliberation as an ideal (Chambers, 2003; Cohen, 1989; Mendelberg, 2002; Gastil, 2008), analysts may wonder if variables have some threshold level that groups must exceed in order to count as being highly deliberative. One of the main research questions for empirical study is: when is group discussion deliberative (i.e., reflective of the norms of deliberative theory)? Additional challenges arise from the complications of studying small group communication, such as the interdependence of data and potential difficulty in getting enough groups together for study to achieve adequate statistical power.

Measures of Deliberation

In the following sections, we discuss studies that have attempted to measure deliberation directly and indirectly. Those that measure deliberation directly examine the deliberative discussion to determine the extent to which the discussion corresponds to theoretical conceptions of deliberation. Studies using indirect measures assess deliberation based on either antecedents (for example, by measuring the extent to which conditions necessary for deliberation are met) or outcomes of the discussion (for example, by measuring post-deliberation changes in participants).

Direct Measurement

The most common direct measurement of deliberation in small groups is what we call discussion analysis, which includes a range of methods used to systematically evaluate the communication engaged in during a deliberative discussion. Another common method used for the direct study of deliberation is to ask participants for their own assessments of the deliberative process. This is typically done through post-deliberation surveys or interviews in which
respondents reflect on their experience as a participant and answer questions about the deliberative quality of the discussion. The case study will be the final direct approach we review.

Discussion Analysis

Discussion analysis involves attempts to directly measure aspects of deliberation by systematically examining aspects of the communication that occurs during a deliberative meeting. This typically involves analyzing records of the interaction as it is preserved in a transcript or video recording of a face-to-face meeting, or in the verbatim, automatically-generated record of an online discussion.

Within the discussion analytic method, there are what we call a micro-analytic versus macro-analytic approach. Micro analysis involves assessing the deliberative quality of discussion discourse through closely analyzing the content of people’s comments during the deliberation. Methods used for a micro-analytic approach include content analysis and discourse analysis. By contrast, the macro-analytic approach asks coders to make summary judgments of the discussion as a whole. We call this a macro-analytic approach because, although it attends to the quality of the deliberative discourse, it is not as closely tied to individual speaking turns and each group member’s contribution to the discussion.

Micro-analytic approaches. An increasingly common approach to studying political deliberations is to analyze the content of the discussions. Krippendorff (2004) defines this type of analysis as “a research technique for making replicable and valid inferences from texts . . . to the contexts of their use” (p. 18). An essential assumption of content analysis is that texts (in this case, deliberations) are meaningful but that such a meaning is often not fixed. In some cases, the researcher is counting the number of times a speaker speaks, or the number of words that are uttered. Such counts are fixed and “objective” in that all reasonable analysts can agree that they
see the same number of words or utterances. Most content analysis, however, focuses on elements of the content where meaning is not fixed. That is, meaning is not found or identified in the manifest features of texts, but rather is interpreted by analysts and by participants in the deliberation. Moreover, context and purpose matters, since meaning is derived from the context in which the communication act occurs (Krippendorff, 2004).

Thus, micro-analysts of deliberations face the daunting task of investigating elements of the deliberation in the face of unfixed or loosely fixed meaning. Put another way, analysts who content analyze deliberations should not enter into the analysis with a view that the content analysis measures “the deliberation,” since there are many aspects to the deliberation, only some of which will be captured by the content analysis that is conducted. Moreover, there are often multiple ways to interpret the same elements within a deliberative event. As a result, content analysis poses challenges to scholars who aim to analyze aspects of the deliberation with an eye towards validity and reliability.

For the content analysis to be reliable, it must be done in such a way that the measurement “responds to the same phenomenon in the same way regardless of the circumstances of its implementation” (Krippendorff, 2004, p. 211). The accepted approach for achieving a reliable measure is first to develop a codebook, which records the sets of definitions and operationalizations of the categories or elements of the deliberation to be analyzed. Second, two or more analysts are trained to analyze the deliberation by applying the rules from the codebook.³ Third, the researcher assesses how much agreement the coders have when analyzing the text. The greater the agreement that exists between the coders, the more reliable the coding

---
³ Increasingly, researchers are trying to automate analysis of texts using computer software. Such efforts have a long history, even within the field of communication (e.g., Roderic Hart’s DICTION, www.dictionsoftware.com), which has even been applied in a deliberative context (Hart & Jarvis, 1999). To date, however, none of these computer-analytic approaches have proven adept at measuring deliberation directly.
scheme is likely to be (for descriptions of content analysis measurement techniques see Krippendorff, 2004, and Neundorf, 2002).

Inter-rater reliability, however, does not guarantee construct validity. Validity in the context of content analysis focuses on how close the measure approximates the theoretical concept. It is possible that two coders can apply the rules of the codebook reliably to an element of a deliberation yet not actually capture the phenomenon of interest. Thus, deliberation analysts must be very careful in crafting the definitions and measurements that comprise the codebook to ensure that they are doing their best to capture accurately the concepts of interest.

In application, content analysis has been used to analyze many different aspects of group deliberation. This wide range includes, for example, the topics raised and the equality of participation (Dutwin, 2003; Gastil, 1993), the balance of opinions (Barabas, 2004), the types of evidence provided by deliberators (Steenbergen, Bachtiger, Sporndli, & Steiner, 2003), argument-based discussion versus bargaining (Holzinger, 2001), the frequency of personal narratives (Dutwin, 2002), the climate of the deliberative group’s opinion (Price, Nir, & Cappella, 2006), and even the overall quality of deliberations (Stromer-Galley, 2007). Other deliberative measures have been developed for analysis of media discourse (e.g., Gerhards, 1997; Page, 1996) but could potentially be adapted to analysis of deliberative groups.

Content analysis schemes that have been developed to assess deliberation typically begin with categories based on political theory and measure the extent to which these categories occur in the actual talk of participants. For example, Hart and Jarvis (1999) analyze the communication that occurred during small group meetings in the 1996 National Issues Convention. Their analytic scheme is based on features of political thought drawn from a number of democratic theorists –coalition formation, integrative complexity, rational action, social constructionism,
and civic republicanism. In their analysis they argue that the pattern derived from these categories, as they are seen in the participants’ discussion, presents an image of participants’ “political model” that characterizes them as distinctively American (p. 64).

There are a few content analytic studies of deliberation that are noteworthy in their presentation of coding schemes that can be used for other deliberative forums. One is Steenbergen et al.’s (2003) measure, the Discourse Quality Index (DQI), and their study of deliberation in British Parliament. The DQI is a coding scheme developed from Habermas’ (1987, 1995) notion of discourse ethics. Steenbergen et al. (2003) argue that high quality (i.e., deliberative) discourse should ideally be open to participation, require assertions to be justified, consider the common good, embody respect for all participants, include constructive politics, and require authenticity or honesty. They translate these theoretical constructs into a content analytic coding scheme to be used with political speeches in parliamentary deliberations. The DQI measures participation, levels of justification given for arguments, the content of justification given, respect for groups of people, respect for counterarguments, respects for others’ demands, and constructive politics. Steenbergen et al. (2003) show that the DQI is a reliable measure, and several of their items (those measuring how people justified their arguments and the extent to which speakers demonstrated respect toward groups of people) hang together and could be used to create a scale to measure the quality of deliberative discourse.

A second noteworthy content analysis study comes from Dutwin (2003), who investigated participation levels within a deliberation and determined whether prior political conversations or political sophistication better prepares participants to engage in deliberation. Dutwin counted the number of times participants spoke and the number of words they produced as well as coded whether participants produced arguments when they spoke. Dutwin developed a
codebook, which included a clear definition and operationalization of “argument.” His trained coders achieved high levels of inter-rater agreement, thereby ensuring a satisfactory level of reliability for each of the coded elements.

The content analysis allowed Dutwin (2003) to analyze whether demographic characteristics such as age, gender, and education levels were correlated with the production of arguments during the deliberation. His results suggested that individuals who were likely to engage in political conversations in their daily lives were more likely to produce arguments during the deliberations. Moreover, he found a fairly balanced amount of participation during the deliberations. Men, for example, did not appear to produce more arguments than women.

Content analysis has its limitations. Perhaps the biggest reason systematic content analysis is not conducted more frequently is the time and labor that it costs. Dutwin (2003) reported that the coding took approximately 100 hours, not counting coder training or the development of the codebook itself. A deliberation content coding project reported by Stromer-Galley (2007) took approximately two months to develop the codebook, an additional 40 hours to train the coder, and then almost six months to do the coding. Part of the reason for such a lengthy process was that 23 groups were included in that project, and each group produced approximately 40 pages of written transcripts. The substantial amount of text that deliberation projects produce can be daunting. For Dutwin’s (2003) project, court reporters were hired to create transcripts of the deliberations on-the-fly. Most projects do not have such financial resources. Instead, projects record the deliberations and subsequently transcribe the results,
which also entails considerable costs. Audio quality problems can make transcription difficult, as can the mere presence of multiple voices on a single audio track.\(^6\)

Moreover, there is risk in putting forth the effort to develop a codebook and train coders, only to discover that the coders cannot get sufficiently acceptable levels of agreement that would serve as the basis for statistical analysis. Published articles rarely include their failures, and we expect that some undiscovered efforts to directly measure deliberation have failed to reach our attention, let alone achieve publication, owing to this difficulty. An unpublished project by the fourth author of this chapter, for instance, attempted to code expressions of disagreement in an online deliberation but found that the coders could not agree as to what counted as an expression of disagreement. Because meaning is not in the text but is constructed in the moment of interaction (Krippendorff, 2004), two reasonable people can construct an interaction and interpret it differently. It is possible, then, that months of time can be spent without any useable data to show for it.

In such a situation, it is possible to shift the analysis away from the systematic counting of elements in a deliberation to a more interpretive thematic approach to analyzing the deliberations. Ryfe (2006), for example, analyzes five National Issues Forums and discovers that a primary form of reasoning is storytelling. He analyzes the role that facilitation plays in promoting narrative as a form of reasoning. He also speculates on why storytelling is such a prominent aspect of the deliberations’ content. Similar works (Black, 2008; Hendriks, 2005; Polletta, 2006; Polletta & Lee, 2006) examine storytelling as important aspects of deliberation and argue that personal stories are important discursive resources that participants draw on during deliberative discussion. Such interpretive analysis, although as uncommon as systematic

\(^6\) For example, The Virtual Agora project (Mühlberger, 2005) that provided the deliberations that Stromer-Galley content analyzed had both a face-to-face and an online condition; due to poor audio quality in the face-to-face groups, transcripts could not be made.
content analysis, is a valid and useful approach to analyzing what comprises a deliberation. It is even more effective when situated as part of a case study, a method discussed later in this chapter.

Another type of micro-level analysis focuses on the interactions and the discursive processes involved in the deliberation. Whereas content analysis measures aspects of participants’ contributions according to a pre-determined coding scheme and typically analyzes data quantitatively, discourse analysts take a more qualitative approach to examining the communication that occurs in a deliberative meeting turn-by-turn. Discourse analytic approaches focus on interaction by examining how people are communicating with each other. Some of the predominant traditions in discourse analysis include conversation analysis, the ethnography of communication, and critical discourse analysis (for an overview of discourse analysis, see Jaworski & Coupland, 1999).

There is a growing body of discourse analytic work in the larger field of political communication that might be relevant to deliberation scholars (e.g., Townsend, 2006; Tracy, 2007), but very little research has been conducted to directly measure aspects of deliberation in this way. One example of a discourse analytic approach to deliberation is Black’s (2006) investigation into how deliberative group members tell and respond to personal stories during disagreements. Black qualitatively analyzed the communication of two groups involved in an online deliberative forum to explore and describe interactive patterns related to storytelling, collective identity, and conflict. She discerned four distinct types of stories, each with a different discursive function. She posited that the different types of stories drew on different ideas about the storyteller’s identity and connection to the group, and that group members use of different types of stories would likely lead to different conflict management strategies.
Some systematic content analysis borrows heavily from discourse analysis to help construct coding categories. Black (2006) uses her findings from the qualitative study of storytelling in online deliberative forums to create a coding scheme to assess the extent to which different types of stories told by participants generated response patterns that led to different conflict management outcomes. Coders were trained to assess aspects of the stories themselves, but also looked at the responses to stories to assess aspects of interaction such as agreement, use of collective identity statements, and conflict management. Similarly, Stromer-Galley (2007) used existing research on expressions of agreement and disagreement in content analysis (see, for example, Pomerantz, 1984). She also approached the coding of the deliberation from the perspective of an interaction. Coders were trained to analyze the deliberation by looking at the thought expressed in the context of what preceded it.

The discourse analytic approach allows researchers to pay close attention to the actual communication that occurred and discern the patterns and meanings that seem most relevant to the case at hand. The drawback to such an approach is that because it is quite time consuming and typically involves paying close attention to a small number of groups, the generalizability of the research findings is limited. However, translating the findings of a discourse analytic study into a content analysis coding scheme allows researchers to examine a much greater number of instances and test hypotheses that could not have been adequately addressed with smaller-scale, qualitative, discourse analytic studies.

**Macro-analytic approaches.** The methods and measures of the studies discussed so far have all operated at a micro-analytic level by focusing on the communication at a rather small unit of analysis (i.e., the speaking turn, argument, thought unit, story, response, etc.). In contrast, a macro-analytic approach to discussion analysis takes a step back to view the conversation as a
whole. In this kind of content analysis, coders are trained to look at the transcript as a whole and make a summary judgment about the quality of the deliberation that occurred. As with micro-level content analysis, coders must be well trained in the concepts and coding rules, and need to reach an acceptable level of inter-rater reliability.

Gastil, Black, and Moscovitz (2008) take this approach by asking coders to read transcripts of group discussions and rate the discussions on six dimensions of deliberation, which they based on the conceptual definition given by Burkhalter et al. (2002). For example, to assess the deliberative dimension of consideration, coders gave each discussion a score between 1 (“strongly disagree”) and 7 (“strongly agree”) in response to the statement, “The group carefully considered what each participant had to say.” In this study, the six dimensions were combined to give each discussion an overall rating of its level of deliberation, as assessed by trained observers. In another example, Black, Welser, DeGroot, and Cosley (2008) operationalize Gastil and Black’s (2008) definition to assess deliberation in some of Wikipedia’s policy-making discussions. To supplement a more micro-level content analytic coding scheme that was used to evaluate each discussion post, coders were trained to read the discussion thread as a whole and rate how well it embodied the analytic and social processes of deliberation.

The macro-level approach does not provide the detailed information about deliberation that is available through micro-level discussion analysis. However, it gives researchers more of a bird’s eye view of the interaction, and may be able to capture important aspects of the interaction that are not evident by studying each individual comment separately. Both of the studies reviewed here use global ratings to supplement other measures, such as micro-level content analysis or self-reports from group discussion participants.
Combining measures allows researchers to examine deliberation in a multifaceted way by providing details of the interaction, a global rating, and the perspectives of the participants themselves. A challenge of a multi-level approach is that different approaches to measurement may yield different conclusions, and results from different measures may not be strongly correlated with one another. As Gastil et al. (2008) note, a lack of association among measures indicates that “a great challenge faces those who hope to operationally define deliberation in a way that is meaningful for participants, practitioners, and researchers alike” (p. ## [will be known soon]).

**Participant Assessments**

Another way to directly assess aspects of deliberation is to ask the participants themselves to comment on their experiences. Rather than closely examining the content of the meetings, participant assessment methods give the deliberators themselves a chance to reflect on their experience with the discussion. Although it is not based on the researcher’s first-hand observation of the meeting, we categorize participant assessment as a direct measure of deliberation because the questions posed to participants are carefully crafted to closely correspond with theoretical definitions of deliberation such as level of respect among participants, relative equality of speaking opportunities, and consideration and comprehension of diverse points of view during the discussion. Researchers typically ask participants to do their assessment through survey questionnaires after completing their discussions (e.g., Halvorsen, 2001; Gastil, Black, Deess, & Leighter, 2008; Reykowski, 2006). In lieu of surveys, in-depth interviews are more typically used as part of a more intensive case study (e.g., Grogan & Gusmano, 2005; Mansbridge, 1980), so the focus here is on questionnaire-based assessments.
One of the most recent efforts to measure deliberation by post-discussion survey is Nabatchi’s (2007) doctoral dissertation on deliberative meetings. Given no consensus definition to work from, she inductively constructed a reliable ten-item scale dubbed the Deliberative Quality Index. Study participants agreed/disagreed with items concerning a range of issues, such as consensus ("Participants worked toward consensus agreements on the issues"), the policies discussed ("A variety of policy alternatives were explored"), values ("The discussions identified shared values in the community"), and consideration ("The discussions helped me consider other sides of the issues"). The items include measures of perceived context (diversity), process (equal speaking turns), and outcomes (changing one’s opinion). In the end, the measure did not prove predictive of changes in participants’ political efficacy, though Nabatchi (1997) speculated that this may have been due to limited variance in her Deliberation Quality Index, an issue we return to in the concluding section of this chapter.

Compared to Nabatchi’s (2007) wide-ranging set of survey items, Gastil (2006) tried to hit on a single, key element of deliberation. His single-item scale measuring participant assessments of public meetings on the activities of Los Alamos Laboratories read,

Sometimes public meetings or group discussions are dominated by people who take up all the meeting time and don't let others speak. In general, were the public meetings you attended dominated by people favorable toward the Lab, dominated by people unfavorable toward the Lab, or did no group dominate the meeting?

A deliberative public meeting was one where participants reported that no group dominated the meeting. Gastil (2006) found that such meeting appeared to yield greater relevant knowledge gains than those meetings that were dominated by one side or the other.
Measures such as these may reach reliability but have limited construct validity as full-fledged measures of the process of group deliberation. Neither focuses exclusively on the quality of the interaction, and Gastil’s (2006) measure would be unable to distinguish a balanced deliberative forum from what was merely a fair fight, albeit an acrimonious and counter-productive affair.

A more promising attempt at participant self-assessment was conducted by Gastil and Sawyer (2004). This unpublished study was unusual in that its primary purpose was to establish the convergent validity of participant and observer assessments of deliberation. The researchers assembled 106 undergraduates into 15 groups, each of which discussed a proposed state law concerning sexual harassment in high schools. Three different measures of deliberation were undertaken: (1) the participants themselves completed a survey including 49 post-discussion assessment items; (2) the forum moderators, all graduate students in a deliberation seminar, completed a six-item assessment after the discussion; and (3) other undergraduates were given 15 minutes of training, then sent to observe the discussion in three-person teams, completing the same six-item assessment aided only by a tally sheet they had used during the discussion to record deliberative moments.

Simplifying the presentation of results in Gastil and Sawyer (2004), three aspects of deliberation were measured using these instruments: the rigorousness of the analytic process (Rigor); the degree to which participants understood each other and considered one another’s arguments (Listening); and the equality of participant speaking opportunities (Equality).\(^7\) For the participant measures, these were three multi-item scales, all with sufficient reliability (minimum alpha = .84), whereas for the moderators and observers, the scales were combinations of their

---

\(^7\) These roughly correspond to the elements of Burkhalter et al.’s (2002) definition of deliberation, and its Listening and Equality elements parallel Reykowski’s (2006) Deliberative Functioning Scale, a multi-item participant assessment instrument.
smaller pool of rating items. All individual scores and ratings were then averaged to create three sets of group-level deliberation ratings—one set for the participants, one for the observers, and one for the forum moderator.

Table 1 shows the associations among these nine ratings. Given the small sample size \( N = 15 \), only a few of the associations were statistically significant, but there were clear indications that for each measurement approach, the different dimensions were positively related. The three participant ratings were moderately correlated (avg. \( r = .37 \)). For both moderator and observer ratings, the Listening scale was modestly correlated with both Rigor and Equality (avg. \( r = .26 \)). Moderators and observers, however, did not have a clear positive association between Rigor and Equality (with the correlation trending negative in the case of moderator ratings).

Comparing deliberative dimensions across the three measurement approaches, other patterns appeared. In the case of Rigor, there was little agreement across methods, with moderators and observers giving positively related Rigor scores \( (r = .17) \) and both having identically negative associations with participants’ self-assessed Rigor scores \( (rs = -.20) \). Moderators and observers were in even stronger agreement on Equality scores \( (r = .56) \) and did not have clear associations with participant ratings. In the case of Listening, however, moderators’ scores had no association with observers or participants, whereas observers and participants had a very high association between their ratings \( (r = .71) \).

This last association is related to the only remaining important correlations across methods and measures. The Listening score from observers’ post-deliberation assessments was predictive of all three participant scores, including the aforementioned correlation, plus average participant scores on Equality \( (r = .72) \) and Rigor \( (r = .44) \).8

---

8 The observers’ Listening score was based on responses to two items: “Discussion participants were able to comprehend the ideas expressed by one another” and “Generally, the group did NOT really consider what
In sum, Gastil and Sawyer’s (2004) study suggests that participants’ assessments of their deliberative experience can differ considerably from equivalent ratings done by forum moderators or trained observers. There is at least some degree of correspondence, however, between participants’ assessments and those of neutral observers from a similar background.  

*Case Study Integration*

The final method that we categorize as a direct measure of deliberation is what we call the integrated case study. The case study utilizes both direct observation of interaction and participant assessments to measure aspects of deliberation that were present in the group. Case studies typically involve prolonged engagement with the group being studied as the researcher engages in participant observation of the group meetings over time.

Mansbridge (1980) provides an early example of an integrated case study of deliberation. Her study of a New England town meetings and a non-profit public health organization provide detailed descriptions of the deliberative processes and the structural conditions that help facilitate deliberation. Like other integrated case studies of deliberation (e.g., Edwards, Hindmarsh, Mercer, Bond, and Rowland, 2008; Gastil, 1993; Grogan & Gusmano, 2005; Karpowitz & Mansbridge, 2005; Mendelberg & Oleske, 2000; Nishizawa, 2005; Polletta, 2002; Wilson, Payne, & Smith, 2003), Mansbridge uses a combination of research methods to provide a rich description of the sites studied and explicate their implications for deliberative theory.

Some recent examples of case studies of deliberation include Ryfe’s (2007) study of a town he calls “Civicville,” Hartz-Karp’s research on forums held in Western Australia (2007), and Polletta’s (2006, 2008, Polletta & Lee, 2006) extensive study of Listening to the City,
AmericaSpeak’s 21st Century Town Hall meeting held in NYC after 9/11. (For more information on this kind of forum see Lukensmeyer, Goldman, & Brigham, 2005.) All these prototypical case studies share a deep engagement with the site of their study, which is partially evidenced by the fact that each followed the groups in question for several years. Another prominent feature of integrated case studies is the use of multiple methods. Ryfe (2007) combines ethnographic observation with in-depth interviews and analysis of newspaper coverage. Polletta’s (2006, 2008) methods include observation, interviews, and content analysis of both face-to-face and online discussions. Gastil’s (1993) involves participant observation, video and transcript analysis, participant questionnaires, and in-depth interviews during which discussants watched a videotaped portion of their meetings to prompt reflection on critical junctures in their deliberations.

Occasionally, deliberation scholars find themselves acting in dual roles, both as a researcher and also a consultant or facilitator for a deliberative group. For instance, in addition to interviewing forum participants after the meetings, Polletta (2008) served on the steering committee for one of the organizations she studied. She also “worked as a facilitator, helped to plan the workshops and to synthesize ideas generated in them for the draft visions, and interviewed organizers” (p. 6). Similarly, Hartz-Karp (2007) served as a consultant to the organization hosting the forums she studied. This dual role of researcher and consultant/facilitator can be tricky because there is potential for conflict between the goals and expectations of these two roles. Hartz-Karp describes herself as “walking a tightrope” between the roles and states, “Naturally it is not possible for me to be completely objective in analyzing a process I had a large share in devising and executing. However, failing to report on the work in which I was involved would constitute a waste of years of experience and hard-won insight” (p.
In these situations, the integrated case study approach can take on a kind of action research paradigm (e.g., Frey & Carragee, 2007) because it involves collaborating with practitioners to address a particular social issue as well as addressing research questions of interest to scholars.

**Indirect Measurement**

The research discussed above presents models for directly measuring aspects of deliberation through focusing close attention to the processes involved in deliberation. In contrast, many scholars measure deliberation indirectly by looking for indicators that deliberation might occur or has occurred. Indirect measures are often used when the antecedents or outcomes of the deliberation are the best (or only) data available to be measured.

For example, in cases of real-world institutionalized small group deliberation, such as congressional committee deliberation, researchers may not have access to the actual deliberation or even to the participants. In such a setting, one may be able to know only if the conditions for deliberation have been satisfied, such that the group is likely to have deliberated. In cases such as jury deliberations, researchers may have access to participants after the deliberative meetings, but not to the content of the meetings themselves.

In other circumstances, researchers may have access to the content of the meetings but choose to focus their analysis on outcomes of deliberation. Deliberation scholars taking this approach can examine the consequences of deliberation and reason that deliberation occurred if participants demonstrate changes that theory claims arise only—or at least principally—from deliberative experiences.

**Antecedents**

Some groups within established governmental institutions routinely claim to have deliberated before making consequential decisions, whether weighing the merits of a proposed
administrative rule change of a resolution to go to war. Frequently, however, such deliberations are not available to researchers to observe, or may be too prohibitive for most researchers to access. Burkhalter (2007) took on the task of examining the conditions for deliberation present in six instances of congressional lawmaking. Some of the conditions for deliberation are enshrined in the U.S. House of Representatives through what is called the “regular order”; however, as Burkhalter found, this regular order is often ignored, along with consideration of diverse views in venues such as legislative hearings. Burkhalter demonstrated that when the majority party’s public message on a piece of legislation was important, the likelihood that deliberation occurred on the legislation declined. The likelihood that deliberation occurred was measured via how much the conditions for deliberation were met in each case.

More commonly, deliberation is simply assumed to have occurred in a sample of public meetings. It is thus a taken-for-granted background variable, actually without any measured variance across individual groups. This is the case in studies that ascribe deliberative discussions to those that take place in the circumstance of a carefully designed, and typically professionally facilitated, public forum. This approach has been taken, for example, in studies of National Issues Forums (e.g., Gastil, 2004; Gastil & Dillard, 1999; Ryfe, 2006) and Deliberative Polls (Fishkin & Luskin, 1999; Luskin et al., 2002; Sturgis et al., 2005). Similarly, Gastil, Deess, Weiser, and Meade’s (2008) study of the civic impact of jury deliberation similarly contrasts the post-service civic experiences of those jurors who had the chance to deliberate in the jury room with those who sat in the jury box during the trial but were dismissed without having the chance to deliberate together. In these cases, the researchers measure outcomes and attribute their presence or absence to the impact of the presumed deliberation.
An oddly similar antecedent approach is that of experimentalists. These researchers set up structural conditions that they believe sufficiently model the presence or absence of one or more deliberative features, then compare the results between deliberative and non-deliberative experimental subgroups. Thus, Morrell (2005) compares an argument-oriented group against one that is agreement-oriented, or non-interactive. Sulkin and Simon (2001) reduce the question even farther, simply labeling an interactive experimental condition as deliberative and compare it against the circumstance under which participants can not interact. In this extreme case, deliberation is simply redefined as merely the opportunity to communicate, surely the minimal structural condition under which deliberation could take place.

More promising is the approach taken by Reykowski (2006), which compared three experimental conditions—minimal conditions for deliberation, introduction of deliberative norms, and support for deliberative norms. In the first condition, the facilitator simply advised the participants at the outset to stay on topic, not offend, and give everyone the chance to speak. In the second, the facilitator introduced a more complete and detailed set of deliberative norms (not unlike those in Burkhalter et al., 2002), which were also given in writing to each participant. The third condition added active intervention by the facilitator, as necessary, to reiterate and reinforce those norms. Unfortunately, post-hoc analysis of the discussions revealed that the experimental conditions were “not very successful in differentiating the conditions of the debates, with all groups showing “some forms of deliberative functioning” (p. 344).

A straightforward drawback to the indirect method of measuring deliberation is that the researcher can only make conclusions regarding the presence or absence of antecedents and cannot speak definitively on whether or not deliberation actually occurred, let alone how it occurred. Nevertheless, in measuring the conditions for deliberation, researchers can make
suggestions regarding aspects of a real-world group’s process that need to be adjusted to make it more likely that the group discussion will be deliberative. This is important when government-sponsored groups that deliberate behind closed doors or at different moments in time regarding the same matter claim that they have deliberated in an effort to provide some legitimacy to the group’s ultimate decision.

Outcomes

An outcome-oriented approach conventionally uses surveys questionnaires, but instead of asking participants to comment on the deliberative process the instruments are designed to measure variables that can reasonably be assumed to have been the result of participating in high-quality deliberations. Mulberger’s (2007) review of the literature shows that deliberation scholars have suggested that post-deliberation, participants could demonstrate increased political knowledge and political reasoning, as well as an increased sense of citizenship. These variables have been measured by a range of studies, which typically use pre- and post-deliberation surveys to compare levels within variables.10

One way to measure political knowledge is to ask participants specific questions about relevant political topics after they engage in group discussions. Jim Fishkin and his colleagues use this approach in Deliberative Polls, by comparing forum participants’ knowledge to the general level of political knowledge in a control group that did not deliberate (Fishkin & Luskin, 1999; Luskin, Fishkin, & Jowell, 2002).

A related measure of political knowledge is what Cappella, Price, and Nir (2002) call “argument repertoire.” These authors argue that “those who can identify multiple explanations with genuine evidence for them, counterarguments to their own explanations, and a resolution in

---

10 Most of these parallel the outcomes hypothesized in Burkhalter et al. (2002), which makes their definition of deliberation particularly apt for considering this parallel set of outcome-based proxy measures for deliberation.
favor of their own explanation are at the highest levels of knowledge about the issue under discussion” (p. 275). Cappella et al. measure participants’ ability to articulate their own opinions about a political topic as well as opinions and arguments given by others. They find that argument repertoire is a good indicator of participation in deliberative conversation, and also that deliberation can lead to an increase in one’s argument repertoire.

Similarly, political attitudes can be measured by asking topical questions about the extent to which participants agree with statements about a relevant topic such as “the United States should continue to engage in military cooperation with other nations to address trouble spots in the world” (Fishkin & Luskin, 1999, p. 24). Studies that investigate attitude change (e.g., Gastil & Dillard, 1999; Gastil, Black, Leighter, & Deess, 2008; Sturgis, Robers, & Allum, 2005) use a pre- and post-forum design to examine the extent to which participating in deliberative forums changes participants’ attitudes about political topics.

Another outcome of deliberation that is often studied is political efficacy, which is the degree to which an individual feels confident that he or she can effectively carry out political action. Craig, Niemei, & Silver (1990) present a measure of efficacy that has been taken up by deliberation scholars to examine the extent to which deliberating with fellow citizens influences one’s level of political efficacy (e.g., Fishkin & Luskin, 1999; Gastil, Black, Deess, & Leighter, 2008; Morrell, 2005).

Some measures used by deliberation scholars that are relevant to a sense of citizenship include interest in politics (Verba, Schlozman, & Brady, 1995), along with political identity, trust in fellow citizens, and trust in governmental institutions (Gastil, Black, Leighter, & Deess, 2008; Muhlberger, 2005, 2007). Survey items developed to measure these variables assess different
aspects of participants’ sense of identity and responsibility as citizens, which can be influenced by participating in deliberative discussion.

Burkhalter et al. (2002) also posit that deliberation can influence participants’ future deliberative habits. One such habit is participating in civic life by voting. In a series of studies, John Gastil and his colleagues (Gastil, Deess, & Weiser, 2002; Gastil, Deess, Weiser, & Meade, 2008) examined the extent to which jury service influenced participants’ future voting behavior. These studies make a noteworthy methodological contribution to deliberative research by combining questionnaire data from jurors with corresponding individual voter-turnout histories, which are publicly-available data. Most deliberative studies that examine outcomes rely solely on survey measures, but the jury studies show that secondary data can be used to supplement primary research techniques.

Indirectly measuring deliberation through survey questionnaires that assess the effects of deliberation has a number of benefits. The methods and measures associated with survey research are well established and surveys can be given to a large number of participants. Such large samples make it possible for researchers to investigate a number of hypotheses related to deliberation and present generalizable conclusions. A liability of this approach is that researchers need to be cautious not to assert that a variable, such as attitude change, is both an indicator of and also caused by deliberation. At the very least, studies relying exclusively on outcome-based measures of deliberation will not be helpful in assessing the impact of deliberation on those same outcomes, lest they be subject to a tautological fallacy.

Summary and Recommendations

Without an agreed upon definition and corresponding measures of deliberation, questions arise regarding whether or not small group deliberation actually achieves the normative
conditions and outcomes that democratic deliberative theory imagines (Levine et al., 2005; Rosenberg, 2005). The lack of agreement on a uniform definition of deliberation from which reliable empirical measures can be derived has already had an impact on the burgeoning field of empirical studies of public deliberation. For example, some empirical observers relying on early conceptualizations of ideal public deliberation articulated by theorists like Habermas (1984) and Cohen (1997) have found actual public deliberations to fall quite short of the normative vision (Rosenberg, 2006; Mendelberg & Oleske, 2000; Mendelberg, 2002; Sanders, 1997; Sturgis et al., 2005; Sulkin & Simon, 2001).

A goal of Gastil and Black’s (2008) project was to distinguish between different types of public deliberation, with group-based deliberation being the best known variant. Further, deliberation researchers should acknowledge that all types of organized small group public deliberation are not alike: It is quite possible that different ostensibly deliberative designs will achieve different levels of deliberation, both within and between different designs. In other words, there is surely a different kind of discussion at a National Issues Forum, as compared to a Deliberative Poll, as compared to a criminal jury or town meeting. Moreover, there is surely considerable variation in the deliberative quality of different events that are all billed as National Issues Forums.

To take just one example, the pitfalls of unstructured discussion have led most deliberation practitioners to conclude that to have a chance at meeting predictions derived from normative democratic deliberative theory, small group deliberation should be structured to include a trained facilitator to orient the group around deliberative norms (Levine et al., 2005). As Ryfe (2006) maintains, facilitation establishes norms and encourages participants to adhere to them. Practitioners who have been involved in overseeing or facilitating deliberative projects
seem to agree that groups need facilitation to reach deliberative outcomes (Gutman, 2007; Mansbridge et. al., 2006). Moreover, Reykowski (2006) demonstrated that the mere presence of a neutral facilitator who articulates even minimal deliberative norms can help groups achieve at least a modicum of deliberation. Even in the midst of this agreement, there are different views of what the facilitator should do to structure the discussion, and there remain those who eschew the idea of facilitation altogether.¹¹

As Gutmann (2007, p. 417) notes, “There are profound differences in the way different [deliberative] initiatives conceive the normative stipulations” of deliberative democratic theory. Fung (2007) urges researchers to try to create uniform conceptual definitions and empirical expectations for theories of democracy because from the standpoint of democratic theory, it is important to test deliberative theory to understand when different practices and institutional designs may promote or undermine democratic principles. In addition, we should avoid holding deliberative democracy to an absolute standard of perfection, something not expected of other regularized forms of political participation (e.g., voting, writing a letter to a member of Congress, expressing one’s opinion at a city council meeting). Grimes (2008) may be right that even “imperfect” processes may deliver the civic benefits theorists ascribe to deliberation, but the larger point is to let go of the ideal altogether when grounding deliberative theory in empirical observation.

All this is to say that the degree of deliberation that occurs in group meetings surely varies, probably to large degrees if we begin studying the wider range of group practices beyond the most carefully constructed deliberative venues and events. If we are to effectively measure

¹¹ For a sense of the range of different deliberative practices, see the National Coalition on Dialogue and Deliberation (www.thataway.org). David Mathews, President of the Kettering Foundation, which begun the National Issues Forums program, has on more than one occasion publicly inveighed against idea of “facilitating” a public forum.
To begin, Table 2 summarizes the pros and cons among the different measurement approaches we have reviewed. The table makes plain that a single approach to measuring deliberation would likely impoverish the field, as each method has a significant pitfall. At the same time, most methods have a compelling advantage that makes them a good candidate for a deliberation measurement toolkit. The lone exception, we believe, is using outcome measures as an indicator of deliberation, a move that renders research unable to test whether those same indicators are, in fact, regular outcomes of deliberation. The research record to date shows that deliberation, like virtually every other civic tonic, delivers some, but not all of the goods the most optimistic theorists have imagined (Delli Carpini et al., 2004; Mendelberg, 2002; Ryfe, 2005).

What is needed is a range of measurement techniques suited to different purposes. On the one hand, we believe the case study approach will always be the best approach for intensive study of specific deliberative events and venues. This is really an umbrella approach, which can incorporate participant observation, in-depth interviews, and even formal content analysis and surveys. As a means of simultaneously making micro- and macro-level observations about deliberative process in all its complexity, we expect the case study approach to be most fruitful as a generative mechanism, pushing forward deliberative theory.

For empirical verification of theory’s claims, larger samples will be required. These, in turn, will make case studies, along with intensive content and discourse analyses, impractical. In these cases, we believe global observer ratings and participant surveys will be necessary. The
survey instruments to date have not been sufficiently road-tested, which will entail validating them against more intensive case study techniques and across a range of deliberative contexts. The Gastil and Sawyer (2004) study illustrates just one part of this process—comparing participant evaluations against those of neutral observers.¹² Much more validation work needs to be done, and publication outlets like the *Journal of Public Deliberation* and *Communication Methods and Measures*, along with the wider community of deliberation scholars, will welcome any such undertaking.

We also wish to stress the potential value of developing short-form versions of any participant assessment tool. Many of the surveys described herein used reliable multi-item scales, but practitioners convening deliberative events may often be willing to sacrifice formal scale reliability to keep their post-discussion survey short. If, for example, a researcher wanted to add regular measures to the surveys produced by the National Issues Forums, it is unlikely that more than a pair of questions could be accommodated, given the existing purposes and logistic constraints of the existing survey. Thus, a long- and short-form version of any measure would be ideal.

At the same time, we believe that researchers should continue to think of deliberation as having distinct dimensions, which we doubt can be reduced to any single-dimensional measure. Given the practical challenges faced by anyone undertaking research on deliberative groups, it may be prudent for researchers to choose to focus their study on the analytic rigor of a deliberative process, on the egalitarian nature of relations among participants, on the degree of respect shown during discussion, or one other element of the process, rather than on all simultaneously. The latter approach sacrifices conceptual and measurement precisions, which is

---

¹² That study also compared participant surveys with forum moderator assessments. In retrospect, it is likely that the cognitive burden of even minimally moderating a forum is too great to expect reliable deliberation ratings from the facilitators themselves.
particularly costly if it turns out that different aspects of deliberation prove to be associated with different deliberative outcomes.

As for the approach that simply inventories the antecedents or structural conditions of deliberation, we believe that particular variants of this approach have value and that this orientation has done the field a favor. The very fact that researchers have carried out fruitful studies by this method should highlight the fact that certain conditions are very conducive to deliberation. This becomes a problem when researchers look for variations in deliberative quality among already well-constructed groups. This problem was noted in Nabatchi’s (2007) study of elaborate deliberative designs and even Reykowski’s (2006) experimental designs. After all, a variable is only useful for testing when one can generate sufficient variance. Those who take the structural antecedent approach to measuring deliberation—and those who assume certain groups were sufficiently deliberative by virtue of their circumstances—appear to have at least some sound basis for their research designs.

Nonetheless, we hope that there is a way of better formalizing the antecedent approach, such that one can assess the deliberative potential of a wide range of groups using similar criteria. One approach might be to assess the degree to which deliberative norms, as articulated by Burkhalter et al. (2002) or others, are articulated by institutional features of the group and by persons with any degree of process authority. Institutional features might range from formal group by-laws to printed posters put on the wall of a meeting venue, and persons with process authority would range from strong facilitators to the members themselves in an egalitarian discussion.

This and the aforementioned approaches should help scholars advance the study of deliberation, and they should also help practitioners systematize their program evaluations.
Working together, deliberative researchers and civic reformers might even use methods like these to bridge the gap between formal academic studies and the more practically-oriented field studies that so rarely appear in scholarly journals. Ultimately, we must bring these strands together if we are to honestly and completely assess the value of public deliberation as a means of advancing the democratic project.
References


Hart, R., & Jarvis, S. (1999). We the people: The contours of lay political discourse. In M. McCombs, & A. Reynolds (Eds.), *The poll with a human face: The National Issues*
Convention experiment in political communication (pp. 59-84). Mahwah, NJ: Lawrence Erlbaum.


*Political Psychology, 27*, 323-346.


Table 1

*Correlations among Observer, Moderator, and Participant Ratings of Analytic Rigor, Careful Listening, and Equality of Speaking Opportunities*

<table>
<thead>
<tr>
<th>Observer Ratings</th>
<th>Moderator Ratings</th>
<th>Participant Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>.17</td>
<td>-.11</td>
</tr>
<tr>
<td>Listening</td>
<td>.09</td>
<td>.04</td>
</tr>
<tr>
<td>Equality</td>
<td>-.02</td>
<td>.31</td>
</tr>
</tbody>
</table>

| Rigor             | .24   | .72**   | -.28   | -.09  | -.02   | .06    | .28    | .40*   |
| Listening         | -.09  | .71**   | .07    | -.36  | -.04   | .30    | .44*   |
| Equality          | .24   | .72**   | -.28   | -.09  | -.02   | .06    | .28    | .40*   |

Note. For one-tailed alpha, ** p < .001, * p < .05, + p < .10. N = 15.
Table 2

*Principal Benefits and Limitations of Different Approaches to Measuring Group Deliberation*

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Principal Benefits</th>
<th>Principal Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Analysis</td>
<td>Close attention to the content of the deliberation.</td>
<td>Time consuming and can be hard to obtain inter-rater reliability.</td>
</tr>
<tr>
<td></td>
<td>Can include large sample.</td>
<td></td>
</tr>
<tr>
<td>Discourse Analysis</td>
<td>Close attention to interaction, patterns that are relevant to the group.</td>
<td>Small sample limits generalizability.</td>
</tr>
<tr>
<td>Global ratings</td>
<td>Integrated assessment of deliberation.</td>
<td>Glosses over distinctions among different facets of deliberation.</td>
</tr>
<tr>
<td>Participant surveys</td>
<td>Systematic assessment by the prospective deliberators themselves</td>
<td>Subject to self-report biases and desire to please event organizers or researchers.</td>
</tr>
<tr>
<td>In-depth participant interviews</td>
<td>Learn about deliberation in participants’ own language</td>
<td>Difficult to compare experiences across individuals and groups.</td>
</tr>
<tr>
<td>Integrated Case Study</td>
<td>Integrates the strengths of different measures.</td>
<td>Time consuming and complicates comparison across different groups.</td>
</tr>
<tr>
<td></td>
<td>Prolonged engagement provides rich detail.</td>
<td></td>
</tr>
<tr>
<td><strong>Indirect Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antecedents/Conditions</td>
<td>Ease of comparison across contexts and wide availability of data.</td>
<td>No knowledge of actual content or process of discussion.</td>
</tr>
<tr>
<td>Consequences</td>
<td>Focuses on those outcomes of deliberation that often originally justified its use.</td>
<td>Unable to assess independently the impact of deliberation.</td>
</tr>
</tbody>
</table>