

11-6-2017

Testing Assumptions in Deliberative Democratic Design: A Preliminary Assessment of the Efficacy of the Participedia Data Archive as an Analytic Tool

John Gastil

Pennsylvania State University, jgastil@psu.edu

Robert C. Richards Jr

Pennsylvania State University, rcr5122@psu.edu


Matt Ryan

University of Southampton, m.g.ryan@soton.ac.uk

Graham Smith

University of Southampton, g.smith@westminster.ac.uk

Follow this and additional works at: <https://www.publicdeliberation.net/jpd>

 Part of the [Comparative Politics Commons](#), [Political Theory Commons](#), [Social Influence and Political Communication Commons](#), and the [Speech and Rhetorical Studies Commons](#)

Recommended Citation

Gastil, John; Richards, Robert C. Jr; Ryan, Matt; and Smith, Graham (2017) "Testing Assumptions in Deliberative Democratic Design: A Preliminary Assessment of the Efficacy of the Participedia Data Archive as an Analytic Tool," *Journal of Public Deliberation*: Vol. 13 : Iss. 2 , Article 1.

Available at: <https://www.publicdeliberation.net/jpd/vol13/iss2/art1>

This Article is brought to you for free and open access by Public Deliberation. It has been accepted for inclusion in Journal of Public Deliberation by an authorized editor of Public Deliberation.

Testing Assumptions in Deliberative Democratic Design: A Preliminary Assessment of the Efficacy of the Participedia Data Archive as an Analytic Tool

Abstract

At smaller social scales, deliberative democratic theory can be restated as an input-process-output model. We advance such a model to formulate hypotheses about how the context and design of a civic engagement process shape the deliberation that takes place therein, as well as the impact of the deliberation on participants and subsequent policymaking. To test those claims, we extract and code case studies from Participedia.net, a research platform that has adopted a self-directed crowd-sourcing strategy to collect data on participatory institutions and deliberative interventions around the world. We explain and confront the challenges faced in coding and analyzing the Participedia cases, which involves managing reliability issues and missing data. In spite of those difficulties, regression analysis of the coded cases shows compelling results, which provide considerable support for our general theoretical model. We conclude with reflections on the implications of our findings for deliberative theory, the design of democratic innovations, and the utility of Participedia as a data archive.

Author Biography

John Gastil is a professor in the Department of Communication Arts & Sciences at The Pennsylvania State University. He received his PhD in communication arts from the University of Wisconsin-Madison in 1994. He specializes in political deliberation and group decision making, and his books include *The Jury and Democracy*, *Democracy in Small Groups* (2nd ed.), *Political Communication and Deliberation*, and *By Popular Demand*.

Robert C. Richards, Jr. is a PhD candidate in the Department of Communication Arts and Sciences at The Pennsylvania State University. He studies legal information and communication systems. His doctoral research concerns the communication of legal information about ballot initiatives to voters. His recent writings have appeared in *Policy & Internet*, *Politics & Society*, and elsewhere.

Matthew Ryan is a Lecturer in Governance and Public Policy at the University of Southampton. His research interests cross the boundaries between political theory and comparative politics, with a strong focus on innovative research methods. His recent research has been published in the journals *Politics* and *Revista Internacional de Sociologia* and in the book *Deliberative Mini-publics: Involving Citizens in the Democratic Process* (ECPR Press, 2014).

Graham Smith is Professor of Politics in the Centre for the Study of Democracy (CSD), Department of Politics and International Relations, University of Westminster. His main research interests are in democratic theory and practice, environmental politics and the politics of the third sector/social economy. His most recent book is *Democratic Innovations: Designing Institutions for Citizen Participation* (Cambridge University Press, 2009), and his recent research has appeared in the journals *Policy & Internet*, *Journal of Political Philosophy*, *Political Studies*, *Environmental Values*, *Ethical Perspectives*, and elsewhere.

Keywords

civic engagement, crowd-sourcing, empirical deliberative theory, public participation

Acknowledgements

We owe a debt of gratitude to everyone on the Participedia.net team for developing the online data archive, as well as to those who have submitted cases to this unique repository. We are also grateful for funding from the University of British Columbia and the Pennsylvania State University to make this research possible. An early version of this paper ("Coding Participedia Cases for Statistical Analysis") was presented at the 2015 Political Studies Association conference in Sheffield, UK.

Experiments with new and traditional modes of public engagement have proliferated in recent years (Warren, 2009). In attempting to make sense of this shift in contemporary governance, democratic theorists, political scientists and participation practitioners have drawn inspiration from deliberative democratic theory (Nabatchi et al., 2012). From this approach, the legitimacy of political decision making rests on the vitality of public deliberation amongst free and equal citizens (Bohman, 1998).

A considerable body of research attempts to analyze the design, process, and consequences of exercises in public engagement from a deliberative perspective, with particular focus on randomly selected mini-publics (e.g., Fishkin, 2009) and participatory budgeting (e.g., Baiocchi, 2005). These designs, however, represent only a small proportion of the diverse universe of democratic innovations. Design features vary considerably among such processes, including the priority given to promoting deliberation amongst participants.

No official records, census, or statistics capture the presence of democratic innovations, let alone the kind of data necessary to test the robustness of assumptions within deliberative democratic theory. Researchers tend to be limited to case studies, often of exemplary cases that skew our expectations of democratic innovations. Larger comparative studies are generally within-type, such as among Deliberative Polls (List et al., 2013), Citizens' Initiative Reviews (Gastil et al., 2016), and participatory budgeting (Sintomer et al., 2012; Wampler, 2007) or within the same political context (Font et al., 2016). Analysis across types and context (geographic and political settings) is relatively rare, since the level of resources required to collect the necessary cases is prohibitive.

The development of Participedia opens up the possibility of such analysis. Participedia (<http://participedia.net>) is a research platform that exploits the power of self-directed crowd-sourcing (Bigham et al., 2015) to collect data on participatory democratic institutions around the world. It is designed explicitly to enable researchers to compare data meaningfully across types and settings, recognizing that such data is held by a diverse group of actors, who organize, sponsor, evaluate, research, or participate in democratic innovations. Participedia has existed since 2009 and currently hosts systematised information on in excess of 650 cases. With the support of a CA\$2.5 million, five-year Partnership Grant from the Social Sciences and Humanities Research Council of Canada (SSHRC), the coverage of cases globally will continue to increase rapidly.¹

This paper exploits the already available data from Participedia to offer the first systematic analysis across a wide variety of political contexts and types of democratic innovations to explore the relationships among design characteristics, deliberative process quality, and impacts on policy and participants. We begin with an account of a stylized input-process-output model intended to capture the relevant core assumptions of deliberative theory. The next section describes the Participedia project and platform in more detail, highlighting how it has been designed to allow the testing of deliberative and participatory theories across a range of cases developed in very different contexts. In the methods section that follows, we explain the challenges faced in coding effectively the Participedia data to accord with our model. This has necessitated not only the use of fixed data from the platform, but also content analysis of case descriptions while overcoming challenges of low levels of inter-coder reliability and missing data. The results show that there are interesting patterns of associations that emerge from the Participedia data. Many of these findings

reinforce existing assumptions about the relationship between design, process and impact, but some may surprise readers and warrant future investigation. We conclude with reflections on the implications of our findings for deliberative theory, our understanding of the design of democratic innovations, and the efficacy of Participedia as a method of generating comparable data in this field of study.

An Input-Process-Output Model of Democratic Deliberation

Drawing an empirical model of the causes and effects of a deliberative process is challenging because the most prominent theories of deliberative democracy emerged out of broader concerns with public spheres and democratic institutions (Chambers, 2003). Those attempts to derive empirical theories have not typically addressed deliberation as embodied in concrete participatory processes (Delli Carpini et al., 2004; Rosenberg, 2007; Warren, 1993). Nonetheless, a subset of the literature has attempted to define and conceptualize deliberative processes in more detail (Burkhalter et al., 2002; Gastil, 2008).

A useful way of organizing the variables implicit in deliberative theories is the input-process-output framework commonly used in small group research (e.g., Pavitt, 1999). This approach separates key concepts into input variables, process variables, and output variables. The inputs, such as a small group's structural features, have effects on the process and outputs but are not themselves subject to change within the theoretical model. The output variables, such as the impact of a group decision, depend on the inputs and process. Finally, the process variables, which include the group's discussion and participants' experience thereof, "mediate" the relationship between inputs and outputs. They are a conduit between inputs and outputs.

An empirical model of a deliberative process can fit within that general framework, provided that one permits leeway for linear causal relationships *within* each of the three broader categories. Figure 1 shows such a model, which will be the starting point for analyzing cases of civic engagement recorded in Participedia. We do not elaborate on this theoretical model in full detail; rather, we seek only to use the model to organize the variables we can extract successfully from the Participedia cases.

The *inputs* within our model are twofold: the context/purpose of the deliberation itself shapes the structural design of the engagement process. For example, a process intended to draft formal legislation, such as the British Columbia Citizens' Assembly (Warren & Pearse, 2008), will be more likely to adopt a deliberative engagement method that emphasizes informational inputs, such as expert witnesses who would testify before a deliberative body.

Structural inputs such as these play a powerful role in determining the democratic and deliberative quality of participant discussions that unfold during public events (Mansbridge et al., 2010; Mendelberg et al., 2014). The *process* variables at the centre of our model encompass two dimensions of democratic deliberation (Gastil, 2008). On the one hand, there is the analytic rigor of the deliberative process, which involves problem analysis, establishing evaluative criteria, and identifying and evaluating solutions. The democratic complement to that is the social relationships established among participants that secure mutual respect, consideration, and equality of opportunity.

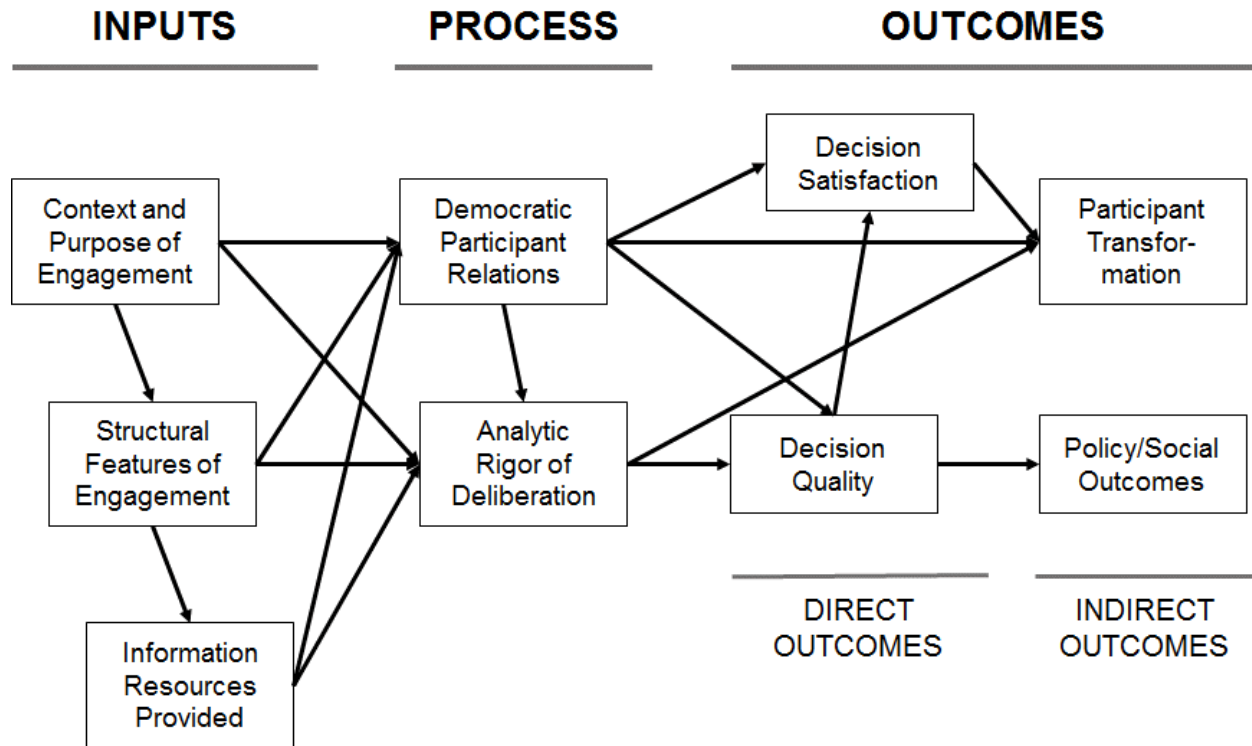


Figure 1. An input-process-outcome model of democratic deliberative civic engagement.

In our model, the democratic social relations are conceptualized as causally prior to analytic rigor, rather than vice versa. This reflects the presumption built into many deliberative processes that active facilitation and democratic “ground rules” must be established to ensure the conditions of mutual respect necessary for the difficult analytic challenges of working through disagreements and trade-offs (Mathews, 1999; Melville et al., 2005; Yankelovich, 1991). The model presumes that the context and design of a deliberative event impinge on both of these process variables.

The *outputs* in this model can be thought of as being near-term and long-term consequences. The most proximate output is the deliberative process’s decision (or recommendation, aggregated opinion, etc.). Both of the process variables are presumed to shape the quality of the decision, which is a basic presumption of epistemic theories of democracy (Landemore, 2013). The quality of the decision, in turn, can have consequences for participants’ decision satisfaction (Gastil et al., 2010).

The more distal impacts are deliberation’s effect on the participants themselves and on public policy, or society more broadly. The first of these is a fundamental assumption of deliberative theory that, on balance, empirical research has supported (Pincock, 2012). Our model presumes that both democratic relations and analytic rigor contribute broadly to changes in participants’ subsequent civic attitudes and behaviour. Moreover, we also expect decision satisfaction to play a role, as was found in the case of the jury’s long-term civic impact (Gastil et al., 2010). As for policy impact, the most common presumption is that it is the quality of decisions made through

deliberative processes that gives them force as a means of shaping public policymaking or broader social processes (Fishkin, 2009; Landemore, 2013).

The input-process-output model that we present and test in this paper has limitations to its empirical validity, which we accept in the name of a practical parsimony. Our model—and the statistical analysis that follows—tests only direct relationships and we avoid the question of interactions and moderator relationships among these variables. Democratic participant relations, for instance, may influence not only decision quality but also the *relationship* between analytic rigor and decision quality. We show nevertheless the value of a more parsimonious empirical test to move forward current understandings of deliberative democracy in practice, while being transparent about any limitations to validity.

The Participedia Case Study Aggregation Project

The Participedia platform provides the first opportunity to draw on data from across types of democratic innovation and political, policy, and geographical settings. It enables researchers to leverage diversity across a large-N of cases. The platform is designed such that researchers, students, practitioners, public officials, and interested publics can upload information on cases of public participation (as well as methods and organizations) in a structured format that allows the cases to be searched and data downloaded. Participedia requests two types of data for cases from contributors. First, contributors are invited to write a free text description describing the key aspects of the particular case. The second type of data is structured: a form requests closed-ended response data about the design and context of the participatory process. This fixed-field data drives the search engine of the platform, allowing users to filter results according to different variables. These results, or all of the structured data for all cases, can be downloaded in a comma-separated values (CSV) file. The wiki-enabled nature of the platform means that users can add, question or revise information. The data collection method for Participedia is thus both structured and decentralized; and the sampling frame is relatively dynamic which creates some challenges but many opportunities for research.

As of March 2017, the platform was home to over 650 cases, with contributions from North America and Europe currently dominating the platform—a function of the location of the most active research teams that have taken the lead in developing Participedia. Though we can have no definitive knowledge of the nature of the population of democratic innovations, it is reasonable to assume that the sample of cases on the platform is somewhat skewed to cases that have proved more effective in either democratic terms or in relation to their impact on the political system. These are the cases that practitioners are more inclined to report, and successful cases are more likely to attract scholars' attention.

That said, Participedia offers significant opportunities for helping us understand the designs of democratic innovations, how they operate, and why they succeed or fail to achieve their objectives. First, since participatory governance is a relatively new area of political practice and study, we are still in the process of clarifying the exact boundaries of the phenomena under investigation. Following J.S. Mill, only through the process of comparison of established and emerging cases can the scope conditions for any modest empirical generalizations for democratic innovations be understood (Mill, 1950). Comparison facilitates both classification and inference. Participedia is uniquely able to support this essential scientific process because the diversity of available cases is

generated by the crowd. Bounded rationalities that dictate what is relevant to this field of study are continuously challenged. Where there has been a conservative tendency to focus on small-N comparisons of most-similar cases, Participedia enables comparison of both more similar and more different cases in order to classify and understand variance.

Second, the forms and context of participatory governance on Participedia are diverse, and any one design involves a trade-off between different democratic qualities (Fung, 2003; Smith, 2009; Thompson, 2008). In other words, the diversity of the sample of democratic innovations on Participedia will include designs with distinct strengths and weaknesses, including many that fail to deliver particular democratic goods. For a field that has been dominated by exemplary case studies and a few small-N comparisons, Participedia represents a significant advance in the diversity of cases, in terms of geography, scale, institutional design, and issue focus. The variety of contributions to the platform ensures that Participedia houses cases with varying democratic attributes, within which public deliberation plays different roles.

Data and Method

The unit of analysis for our study is an individual article (or “case”) in Participedia. Our aim was to explore associations among key features of participatory processes. However, a fundamental methodological challenge was to draw a sample from a platform whose sampling frame is dynamic and crowd-sourced and informed by a field of study with no definitive knowledge of the population of democratic innovations.

To address these challenges, we took a two-step approach. First, we aimed to generate variability in the deliberative characteristics of cases in order that we could study the relationships between such variations within our input-process-output model. To ensure diversity of cases, including those not designed to embody deliberative practices, we applied a sampling matrix using three fixed-field variables that are requested from contributors and that have been shown to have theoretical and empirical significance from previous studies of the deliberative qualities of democratic innovations. These are the presence or absence of facilitation, the use of active or passive modes of interaction among participants, and the application of voting or some other decision method.

Second, the crowd produces data of varying quality. Since we needed to develop variables from the text description, a decision was made for each case as to whether the description was extensive enough to enable effective content analysis.

This purposive sampling procedure resulted in 167 cases selected from the 304 cases that at the time of analysis made up the Participedia dataset. Of the 167 cases coded, those cases with the lowest inter-rater reliability were omitted (as discussed below), resulting in a sample of 149 cases.

In this final sample, 66 percent of cases were facilitated. Active interaction modes were employed in 58 percent of cases, passive modes in 17 percent, and both active and passive interactive modes in 24 percent. In terms of decision method, 33 percent of cases employed voting exclusively, 35 percent a non-voting method, 2 percent used multiple decision methods, in 19 percent participants took no decision, and the decision method was unknown in 11 percent of cases.² As for the

geographic distribution of the final sample, 42 percent of events took place in North America, 35 percent in Europe, 10 percent in Asian nations, 9 percent in South America, and 3 percent in Africa.

The sheer variety of designs that the sample incorporates can be seen by contrasting Australia's First Citizens' Parliament (held in the Old Parliament House in Canberra), a randomly-selected mini-public, and a Rural Plebiscite Experiment in Indonesia. The sampling matrix captures the similarity between the two initiatives in that they both applied voting procedures to make decisions. They differ, however, in that the former was facilitated and involves active interaction between participants; the latter is not facilitated and had no formal interaction process. Other variations abound in this dataset, which includes a Brazilian Municipal Health Council, participatory budgeting processes of varied forms, virtual forums, town meetings, and more.

From these cases, we extracted variables to represent each of the nine categories that make up the input-process-output model. Table 1 provides a summary of these variables in relation to our theoretical model. It provides information on the number of items that make up the variable (including α score), the number of cases constituting each item, the variable scale, and its mean. Only five of the variables are taken directly from the fixed-field data, with the rest being the result of coding. In the remainder of this section, we provide an overview of the coding process used to extract those variables from the Participedia data set, as well as our statistical approach to analyzing them. The online appendix accompanying this article³ provides additional information on item wording and inter-item scale reliability.

Content-Analytic Method

The case content analyses conducted by one of the authors and undergraduate research assistants generated the majority of the variables. Content analysis used a codebook that operationalized the variables set out in Gastil, Knobloch, and Kelly's (2012) framework for analyzing and evaluating participatory democratic events. The codebook asked coders to make objective, not relative, assessments.⁴

Content analyses were conducted using Neuendorf's (2002) "descriptive" coding method (pp. 53–54). Inter-rater reliability was measured by means of Pearson correlation coefficients. When codings by all 20 coders are included, correlations between coders were low (raw average correlation: $r = .16$; raw average correlation with original [expert] coder: $r = .17$; weighted average correlation: $r = .18$; weighted average correlation with original [expert] coder: $r = .14$). When codings of the three coders with the lowest correlations with other coders were omitted, correlations between coders increased slightly but remained low (Figure 2) (raw average correlation: $r = .20$; raw average correlation with original [expert] coder: $r = .28$; weighted average correlation: $r = 0.206$; weighted average correlation with original [expert] coder: $r = .18$).

Table 1. *Variables from Participedia Fixed-Field Data and Coded Case Narratives*

	Variable	Description (Fixed-Field Data Noted)	Items (<i>a</i>)	<i>N</i>* (Lo-Hi)	Scale	Mean (SD)
Input: Context	Consultation	Purpose of event was a consultation (fixed-field)	1	149	0-1	.48 (.5)
Input: Structural Features of Engagement	Random Sample	Random sample used (fixed-field)	1	143	0-1	.34 (.47)
	Representative Sample	Sample was representative	1	116	1-5	3.43 (.81)
	Stakeholder Role in Design	Sufficiency of the role stakeholders had in the design process	2 (.59)	79 (93-113)	1-5	3.42 (.76)
	Interaction	Level of interactivity (passive=0, mixed=1, active=2) (fixed-field)	1	149	Nominal	--
	Facilitated	Process was facilitated (fixed-field)	1	149	0-1	.66 (.48)
	Deliberative Process Clarity	Trained facilitators led a process with a clear charge and ground rules for respect	5 (.84)	110 (36-141)	1-5	3.86 (.68)
	Small Group Dialogue	Participants were encouraged to have small group dialogues	3 (.75)	103 (42-98)	1-5	3.3 (1.01)
	Voting	Process involved voting (fixed-field)	1	140	0-1	.42 (.5)
Input: Information Resources Provided	Materials Quality	Quality of materials provided for discussion	9 (.88)	93 (56-94)	1-5	3.61 (.57)
	Knowledgeable Witnesses	Witnesses testifying were knowledgeable	4 (.76)	91 (54-113)	1-5	3.87 (.64)
	Witnesses Diverse and Civil	Witnesses were selected carefully to be diverse and civil	5 (.84)	69 (25-68)	1-5	3.6 (.75)
	Witness Time	Witnesses were allotted sufficient time to testify and be cross-examined	2 (.86)	49 (37-53)	1-5	3.6 (.87)
Process: Relations	Democratic consideration	Panel was reflective, considered different points of view, and embraced disagreement	6 (.86)	88 (55-89)	1-5	3.62 (.53)
	Democratic respect	Participants were respectful and careful about giving time to each other and to different views	8 (.88)	55 (24-55)	1-5	3.74 (.69)
Process: Analysis	Deliberative Breadth	Deliberation had strong breadth of knowledge, values, and solutions	11 (.93)	125 (31-118)	1-5	3.64 (.55)
	Deliberative Depth	Deliberation had time for in-depth questions and reflection	9 (.90)	110 (41-110)	1-5	3.42 (.59)
Direct Outcomes	Decision Quality	The process' final report/statement/decision was accurate, well-reasoned, and comprehensive	6 (.75)	77 (52-75)	1-5	3.76 (.53)
	Participant Decision Rating	Participants rated decision as satisfying and are committed to it	2 (.55)	66 (76-85)	1-5	3.6 (.88)
Indirect Outcomes	Policy Influence	Process influenced government or organizational policymaking	2 (.72)	127 (104-159)	1-5	3.54 (.93)
	Change in Participants' Civic Engagement	Process increased panelist civic engagement in short-term	7 (.89)	63 (29-50)	1-5	3.75 (.58)

Note. * This field shows the *N* achieved for this variable by using a scale creation equation that ignores missing values, with the Lo being the single item with the fewest cases and the Hi being the item with the most cases. By averaging items (and ignoring missing cases), a larger *N* for the scale emerges.

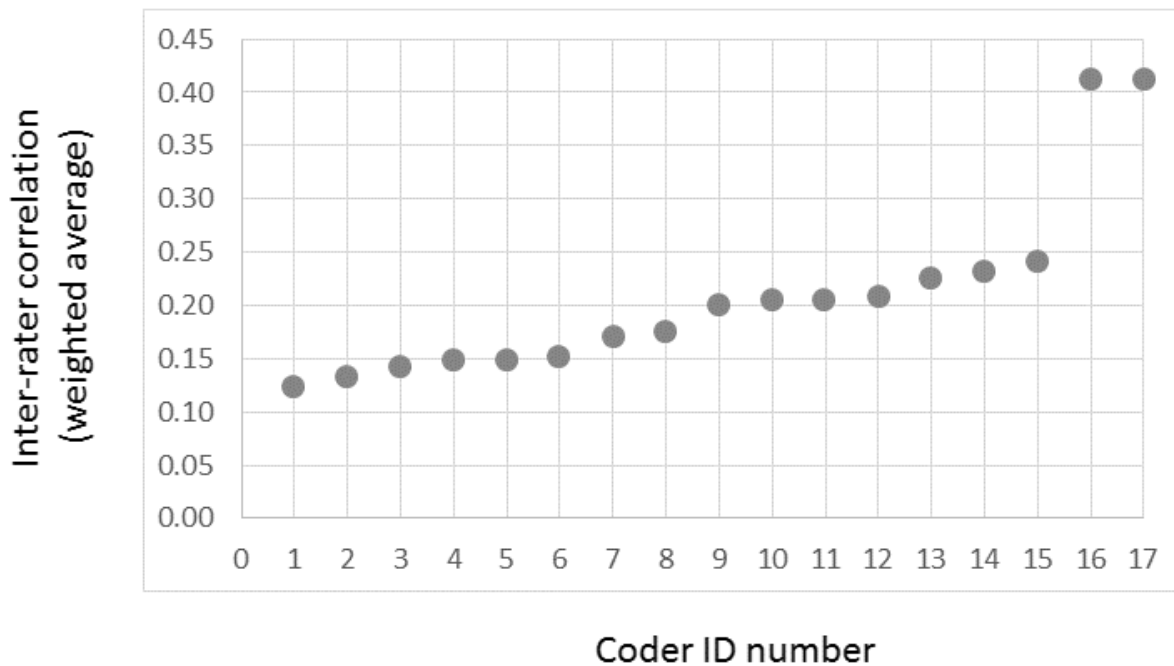


Figure 2. Inter-rater correlations (weighted average), with weakest three coders omitted.

Note. $N = 3499$ codings, 17 coders. Y axis shows the mean Pearson correlation coefficients among coders' codings.

To construct each item, codings were averaged across all coders who coded the item, and missing codes were ignored. Coded values indicating that the code was inapplicable to the case or that the coder was unable to determine the appropriate value for the variable from the text of the Participedia case article were treated as missing values. Next, conceptually similar items were identified for possible aggregation into multi-item scales, and the inter-item scale reliability of each such group of items was measured using Cronbach's *alpha* (Cronbach, 1951; DeVellis, 2012). Notwithstanding low inter-rater reliability, inter-item scale reliabilities (Table 1, and Tables 1A and 2A in the Appendix) were generally acceptable (Cronbach's α : median = .85, mean = .81) (Rosenthal & Rosnow, 2008).

Low inter-rater reliability is not necessarily inconsistent with sufficient inter-item reliability or with adequate construct validity. As Haskard and colleagues (2009) observe, when inter-rater correlations are low, coders may nonetheless be identifying "different but complementary aspects of the variable they are rating" (p. 26). As a result, when ratings that are poorly correlated with each other are aggregated into items relating to a common factor, those items may prove adequately correlated with each other, and thus yield sufficient scale reliability. Scale reliability is, in turn, a measure of construct validity (Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Rosenthal & Rosnow, 2008). The coincidence of low inter-rater reliability with adequate inter-item scale reliability resembles other types of statistical amalgamation paradoxes (Good & Mittal, 1987),

such as Simpson's paradox (Samuels, 1993), in which patterns observed among data at a granular level of analysis are substantially different or even reversed at a higher level of analysis.

Once adequate inter-item scale reliability had been established, conceptually related items were combined into multi-item scales by summing the means of related items and ignoring missing values. Amalgamating conceptually related items enabled the researchers to take into account the inapplicability of particular items to particular cases while maximizing the utility of available items relating to the same concept.

For example, for the multi-item scale *Deliberative Process Clarity*, of the five items making up the scale, three had fewer than 100 valid cases out of a sample of 149 cases ("Rules for talk were sufficiently explained to the panelists": $n = 56$; "Rules for talk required panelists to treat each other with respect": $n = 36$; "Staff modeled respectful behavior": $n = 39$), and one item ("The charge or question was clearly put to the panelists") was more weakly correlated with the others (i.e., r values ranging from .32 to .47). When the items were combined in a single, multi-item scale, the number of valid cases rose to 110, and good inter-item scale reliability was achieved (Cronbach's $\alpha = .84$). (See Table 1, and Tables 1A and 2A in the Appendix.) Thus, aggregating conceptually similar items in multi-item scales enabled the inclusion in the analysis of related measures of core concepts while rendering usable items having relatively few valid cases.

After multi-item scales had been constructed, missing values for those scales were imputed via multiple imputation (Graham, 2012). The frequency distribution of the mean of all 16 multi-item scales plus the sole single-item scale (*Representative Sample*) approximated a normal distribution, as shown in Figure 3 and Table 2. Also, these scales were moderately and positively correlated with each other. (Mean pairwise correlation among scales was $r = .23$.)

Method of Statistical Analysis

With one exception (described below), data were analysed using bootstrap regression,⁵ with each equation including those groups of independent variables theorized as being proximate causes of each successive dependent variable in Figure 1. The main statistical model presented addresses missing data by using listwise case deletion. In addition, observations with outlier residuals were omitted before analysis.⁶ At the end of the Results section, we review the findings of alternative models to demonstrate the consistency of findings across different models.

The principal cause of reduced sample size, however, was neither listwise deletion nor outlier removal, but rather the difficulty of coding many cases owing to insufficiently detailed descriptions in the relevant Participedia case. We revisit this challenge in our concluding section, when we suggest the implications of this study for Participedia and similar data repositories.

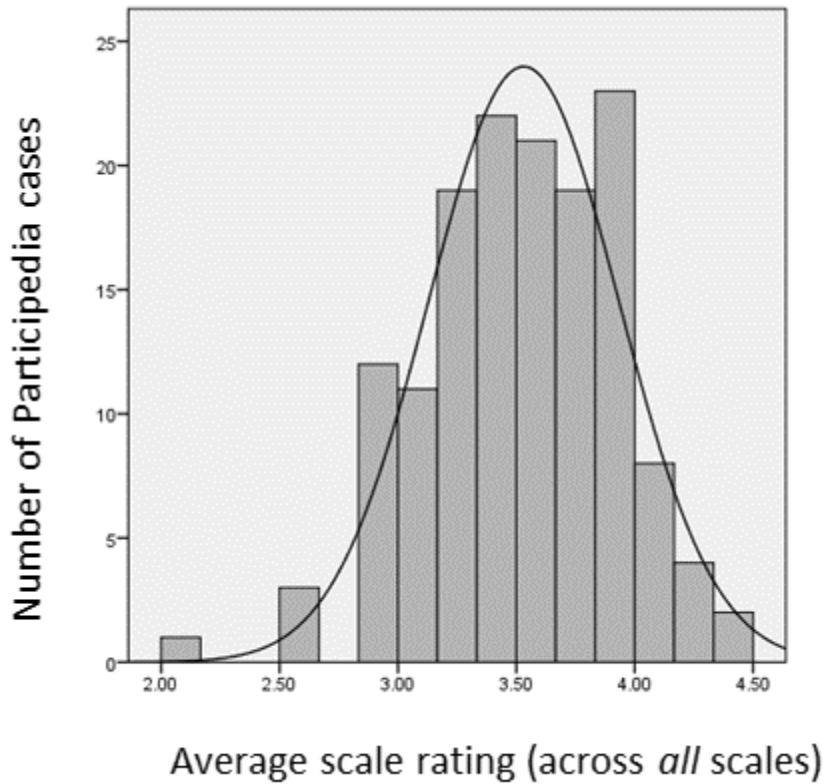


Figure 3. Frequency distribution of average scale rating across all scales.

Note. $N = 145$. Average scale rating is the arithmetic mean of the 16 multi-item scales and the single-item scale (*Representative sample*) employed in the analysis. See Tables 1, 1A, and 2A, and text for details.

Results

Relationships among Input Variables

Relations among input, process, and outcome variables (described in Table 1) were assessed in a series of statistical tests, the first of which were simple t-tests and chi-square analyses of the relationship between the context/purpose of engagement (i.e., the dichotomous *Consultation* variable) and the eight variables measuring the structural features of engagement. Only one of these relationships reached significance, with voting occurring in only 28 percent of the cases of policymakers consulting citizens, compared to its use in 54 percent of all other cases, $\chi^2_{(1, N=140)} = 9.50, p = .002$.

Table 2

Descriptive Statistics: Average Scale Rating Across All Scales

Mean	3.53
Median	3.55
Mode	3.00
SD	.40
Minimum	2.00
Maximum	4.48
	10th 3.00
	20th 3.19
	30th 3.33
	40th 3.42
Percentiles:	50th 3.55
	60th 3.64
	70th 3.78
	80th 3.93
	90th 4.00
Skewness	-.37
Kurtosis	.58

Note. $N = 145$. Entries are descriptive statistics for the arithmetic mean of the 16 multi-item scales and the single-item scale (*Representative Sample*) employed in the analysis. See Figure 3, Tables 1A and 2A, and text for details.

The next set of hypothesized associations shown in Figure 1 were from the context/purpose and structural features of engagement to the information resources provided to participants, as measured by four variables. Table 3 summarizes the results of these bootstrap regression models. Associations were scattered across the models, with the most consistent variable being *Stakeholder Role in Design*, which was positively and significantly associated with higher-quality materials and more diverse and civil witness participation; *Stakeholder Role* also had associations approaching significance⁷ with the inclusion of more knowledgeable witnesses, but also less total time allocated to witnesses. The only variable having a positive and significant association with the use of a more knowledgeable set of witnesses was the presence of a relatively representative sample of participants. Further, *Deliberative Process Clarity* was positively and significantly associated with the use of diverse and civil witnesses, and passive interaction had a negative and significant association with time allocated to witnesses.

Table 3

Bootstrap Regression Models of Information Resources in Participedia Cases

Variable	Materials Quality	Knowledgeable Witnesses	Witnesses Diverse/Civil	Witness Time
Input: Context				
Intended purpose: Consultation	.10 (.18)	-.02 (.14)	-.21 (.19)	-.05 (.32)
Input: Structural Features				
Random Sample	.01 (.16)	-.09 (.15)	.11 (.18)	-.31 (.39)
Representative Sample	.15 (.11)	.31* (.13)	-.03 (.11)	.39 (.26)
Stakeholder Role in Design	.25* (.12)	.16† (.09)	.31* (.14)	-.56† (.28)
Interaction (Active Only)	.16 (.16)	.05 (.12)	-.11 (.17)	-.62 (.40)
Interaction (Passive Only)	-.25 (.24)	-.29 (.34)	-.14 (.29)	-1.14* (.55)
Facilitated	-.19 (.21)	.27 (.21)	-.35 (.24)	-.08 (.38)
Deliberative Process Clarity	.27 (.16)	.22 (.15)	.48** (.16)	-.25 (.32)
Small Group Dialogue	.04 (.10)	.03 (.07)	.12 (.10)	.13 (.17)
Voting	.10 (.17)	-.24 (.16)	.15 (.20)	-.61 (.42)
Constant	1.20 (.75)	1.26† (.67)	.78 (.74)	5.87** (1.40)
Observations	51	52	47	35
Number of bootstrap samples	1119	1116	1116	1114

Note. Cell entries are estimated unstandardized bootstrap regression coefficients. Parentheses indicate bootstrap standard errors. Bold type indicates a significant coefficient for a variable in the model (* $p < .05$, ** $p < .01$, in both instances two-tailed). Dagger indicates a coefficient approaching significance for a variable in the model († $p < .10$, two-tailed). Missing data are addressed through listwise deletion. Observations with outlier residuals (i.e., Cook's D exceeds cutoff stated in Belsley et al. [1980] and t test corresponding to the Studentized residual has $p < .05$, two-tailed; or COVRATIO exceeds the cutoff stated in Belsley et al. [1980]) were omitted before analysis. Bootstrapping algorithm was configured for 1120 samples, with 95-percent confidence intervals estimated using the bias-corrected and accelerated method. In some samples, one or more variables were constant or had missing correlations, causing some estimates to be based on fewer than 1120 bootstrap samples. Method: Bootstrap regression in SPSS.

Relationships Between Input and Process Variables

Associations were estimated between structural features and information resources (two types of input variables) and the two process variable categories (democratic relations and analytic rigor). In addition, associations were estimated between democratic relations and analytic rigor, since theory suggests that the former enable the latter. The findings for all these models appear in Table 4.

The extent to which deliberative events described in Participedia exhibited a democratic social process was addressed in the bootstrap regression models for *Democratic Consideration* and *Democratic Respect* shown in Table 4. In the former model, *Materials Quality* and *Witness Time* were positively and significant associated with *Democratic Consideration*, whereas *Deliberative Process Clarity* was negatively and significantly associated with the response variable. The model of *Democratic Respect* yielded one expected coefficient approaching significance, and one unanticipated significant coefficient. The *Deliberative Process Clarity* variable had a positive association with *Democratic Respect* that approached significance, but *Facilitation*'s association was negative and significant—an unexpected finding we discuss later.

Relationships between inputs and democratic-social interaction, on the one hand, and the analytic rigor of a process, on the other, also varied between *Deliberative Breadth* and *Deliberative Depth*. As shown in Table 4, *Witness Time* was positively and significantly associated with *Deliberative Breadth*, and *Deliberative Process Clarity* had a positive association approaching significance with that response variable ($p = .08$).⁸ In the model of *Deliberative Depth*, *Democratic Consideration* had a positive and significant association with the response variable,⁹ while the association between *Small Group Dialogue* and the response variable was positive and approached significance ($p = .09$).

Relationships Among Process and Outcome Variables

The limited number of direct paths to output variables in our theoretical model makes summarizing the final regression equations relatively straightforward (Table 5). Consistent with our theorizing that *Decision Quality* would partly be a function of the democratic-social dimension of the deliberative process, *Democratic Respect* was positively and significantly associated with *Decision Quality* and the relationship between *Democratic Consideration* and *Decision Quality* was positive and approached significance ($p = .07$).

In the model of *Participant Decision Rating*, the lone significant association was a positive one between *Deliberative Breadth* and the response variable. The model of *Change in Participants' Civic Engagement* yielded two significant and positive associations, with *Decision Quality* and *Deliberative Breadth*.

A final model of *Policy Influence* had mixed findings, some of which ran contrary to expectations. *Democratic Consideration* had a strong positive relationship with *Policy Influence*, and *Deliberative Breadth* had a positive association with the response variable that approached significance ($p = .09$). There was also a negative and significant association for *Deliberative Depth*, while *Democratic Respect* had a negative and nearly significant relationship with *Policy Influence* ($p = .06$).

Table 4

Bootstrap Regression Models of Process Measures in Participedia Cases

Variable	Democratic Consideration	Democratic Respect	Deliberative Breadth	Deliberative Depth
Input: Structural Features				
Random Sample	--	--	--	--
Representative Sample	--	--	--	--
Stakeholder Role in Design	--	--	--	--
Interaction (Active Only)	--	--	--	--
Interaction (Passive Only)	--	--	--	--
Facilitated	.13 (.14)	-.64* (.30)	-.32 (.25)	-.00 (.32)
Deliberative Process Clarity	-.22* (.10)	.45† (.23)	.27† (.18)	-.07 (.20)
Small Group Dialogue	.06 (.07)	.12 (.14)	.00 (.10)	.21† (.13)
Voting	--	--	--	--
Input: Information				
Materials Quality	.40* (.18)	.18 (.29)	.33 (.26)	.17 (.30)
Knowledgeable Witnesses	.21 (.17)	.36 (.35)	.31 (.34)	.07 (.25)
Witnesses Diverse and Civil	.09 (.10)	-.04 (.18)	-.12 (.16)	.02 (.17)
Witness Time	.19* (.08)	.13 (.14)	.21* (.10)	.06 (.14)
Process: Dem. Relations				
Democratic consideration	--	--	.00 (.29)	.57*# (.26)
Democratic respect	--	--	.02 (.16)	-.08 (.18)
Constant	1.00† (.59)	-.23 (.90)	.06 (.71)	.26 (.84)
Observations	32	32	29	30
Number of bootstrap samples	1120	1119	1115	1119

Note. Cell entries are estimated unstandardized bootstrap regression coefficients. Parentheses indicate bootstrap standard errors. Bold type indicates a significant coefficient for a variable in the model (* $p < .05$, ** $p < .01$, in both instances two-tailed). Dagger indicates a coefficient approaching significance for a variable in the model († $p < .10$, two-tailed). Pound sign (#) indicates that bias-corrected confidence intervals include zero, whereas confidence intervals estimated by the percentile method exclude zero. Missing data are addressed through listwise deletion. Observations with outlier residuals (i.e., Cook's D exceeds cutoff stated in Belsley et al. [1980] and t test corresponding to the Studentized residual has $p < .05$, two-tailed; or COVRATIO exceeds the cutoff stated in Belsley et al. [1980]) were omitted before analysis. Bootstrapping algorithm was configured for 1120 samples, with 95-percent confidence intervals estimated using the bias-corrected and accelerated method. In some samples, one or more variables were constant or had missing correlations, causing some estimates to be based on fewer than 1120 bootstrap samples. Method: Bootstrap regression in SPSS.

Table 5

Bootstrap Regression Models of Outcome Measures in Participedia Cases

Variable	Decision Quality	Participant Decision Rating	Influence on Policy	Change in Participants' Civic Engagement (Short-Term)
Process: Democratic Relations				
	.28† (.16)	.23 (.23)	.94* (.44)	.10 (.17)
Democratic consideration				
	.21* (.10)	.20 (.22)	-.61† (.32)	-.08 (.18)
Democratic respect				
Process: Analytic Rigor				
	-.03 (.16)	.68** (.22)	.64† (.39)	.83** (.21)
Deliberative breadth				
	.05 (.11)	-.07 (.21)	-.72* (.35)	-.14 (.13)
Deliberative depth				
Outcomes: Direct				
	--	-.22 (.24)	.52 (.47)	.36* (.17)
Decision quality				
	--	--	.12 (.19)	-.08 (.17)
Participant decision rating				
	1.90** (.42)	.58 (.75)	-.09 (1.53)	-.12 (.62)
Constant				
Observations	44	40	41	30
Number of bootstrap samples	1120	1120	1120	1120

Note. Cell entries are estimated unstandardized bootstrap regression coefficients. Parentheses indicate bootstrap standard errors. Bold type indicates a significant coefficient for a variable in the model (* $p < .05$, ** $p < .01$, in both instances two-tailed). Dagger indicates a coefficient approaching significance for a variable in the model († $p < .10$, two-tailed). Missing data are addressed through listwise deletion. Observations with outlier residuals were omitted before analysis. For the model of Influence on Policy, “outlier residual” means Cook’s D exceeded cutoff stated in Belsley et al. (1980) and t test corresponding to the Studentized residual had $p < .05$, two-tailed; or COVRATIO exceeded the cutoff stated in Belsley et al. (1980). For the other three models, “outlier residual” means Cook’s D exceeded cutoff stated in Belsley et al. (1980), or COVRATIO exceeded the cutoff stated in Belsley et al. (1980). Bootstrapping algorithm was configured for 1120 samples, with 95-percent confidence intervals estimated using the bias-corrected and accelerated method. Method: Bootstrap regression in SPSS.

Alternative Model Considerations

To test the robustness of the results reported in Tables 3, 4, and 5, we ran a series of alternative bootstrap regression models (see Tables 3A through 14A in the appendix). These models varied in terms of whether variables with non-normal frequency distributions were transformed,¹⁰ whether variables were centered and standardized, whether missing data were addressed through pairwise or listwise deletion, whether observations having outlier residuals were retained in the dataset, and how “outlier residuals” were defined. (For details, see the “General Note for Tables 3A through 15.2A” on page 8 of the Appendix.)

All models whose results are reported in Table 3, 4, or 5 had at least some results partially corroborated by results of alternative models.¹¹ The strongest corroboration was found for the models of *Deliberative Depth* and *Participant Decision Rating*, all of whose significant associations were fully corroborated in one or more alternative models (see Tables 10A and 12A in the Appendix). The weakest corroboration occurred with respect to models of *Materials Quality*, *Democratic Consideration*, and *Decision Quality*, almost all of whose significant associations only approached significance in alternative models (Appendix Tables 3A, 7A, and 11A).¹² For the remaining response variables, alternative models fully corroborated some results from Table 3, 4, or 5, and partially corroborated others.

Of the independent variables analysed in this study, *Stakeholders' Role in Design Process* and *Deliberative Breadth* had results that were most often fully corroborated in alternative models. Each of these independent variables had coefficients in two of the models reported in Table 3, 4, or 5 that were fully corroborated by two or more alternative models (for *Stakeholders' Role in Design Process* these were models of *Knowledgeable Witnesses* and *Diverse and Civil Witnesses*, and for *Deliberative Breadth* these were the models of *Participant Decision Rating* and *Change in Participants' Civic Engagement [Short-Term]*) (Appendix Tables 4A, 5A, 12A, and 14A). In addition, each of these independent variables had a result that was fully corroborated in one alternative model (for *Stakeholders' Role in Design Process* this was the association with *Witness Time*, and for *Deliberative Breadth* this was the association with *Policy Influence*) (Appendix Tables 6A, 13.1A, and 13.2A).¹³ Further, the independent variable *Small Group Dialogue*'s nearly significant and positive association with *Deliberative Depth* was corroborated in two alternative models (Appendix Table 10A).

One independent variable—*Representative Sample*—had no results corroborated by alternative models. In addition, three of the significant or nearly significant associations involving the independent variable *Deliberative Process Clarity* (in the models of *Diverse and Civil Witnesses*, *Democratic Consideration*, and *Deliberative Breadth*) were not corroborated in any alternative models, although *Deliberative Process Clarity* had one association approaching significance with *Democratic Respect* that was fully corroborated in three alternative models (Appendix Tables 5A, 7A, 9A, and 8A). Moreover, for five independent variables (*Interaction [Passive-Only]*, *Facilitated*, *Materials Quality*, *Deliberative Depth*, and *Decision Quality*), only partial corroboration of significant results was yielded by alternative models. For each of the remaining independent variables, alternative models fully corroborated some results reported in Table 3, 4, or 5 and partially corroborated others.

Results of alternative models also shed light on the unexpected results reported in Tables 4 and 5. The negative association between *Facilitated* and *Democratic Respect* was only partially corroborated in one alternative model. Likewise, the negative association of *Deliberative Depth* with *Policy Influence* was partially corroborated in two alternative models, and the negative association approaching significance between *Democratic Respect* and *Policy Influence* was fully corroborated in two alternative models (Appendix Tables 8A and 13.2A).

Discussion

This analysis of Participedia case information aimed to furnish new insights into associations between the context, design attributes, deliberative processes, and impact of participatory

processes. Results of this analysis disclose important relationships among such variables, and they allow us to explore different potential pathways within our model. In the summary that follows, we focus on those statistical relationships that showed statistical significance most consistently across the alternative regression models we constructed.

Turning first to design inputs, the importance of the role of stakeholders is clear and holds across models: stakeholder involvement in design had multiple connections to information resources. The presence of stakeholders suggests a desire to ensure that their perspectives on the issue at hand are understood by participants (Kahane et al., 2013). The negative association with the time given to witnesses is intriguing. Potential explanations include attempts to simplify logistics and reduce costs associated with inviting witnesses and reduction in more neutral witnesses following complaints from key stakeholders in some deliberative processes (Kahane et al., 2013, p. 21, note 20). This occurred with the 2014 iterations of the Oregon Citizens' Initiative Review, which initiated those changes to simplify logistics and reduce costs (see Gastil et al., 2015).¹⁴

Aspects of information resources—the quality of informational materials and the time provided for witness-participant interaction—employed in democratic innovations were found to be significantly associated with participants' consideration of each other's arguments. Such a finding accords with theories that stress the role of a high quality "information base" in fostering democratic deliberation (Gastil, 2008, p. 9; Fishkin, 2009). The time afforded to witnesses to testify and respond to participants' questions was also positively associated with the comprehensiveness of participants' deliberative analysis of issues and policy solutions, and corroborated in one alternative model.

In addition, the clarity of the deliberative process design was positively associated with respectful behaviour among participants and witnesses, in accord with previous deliberative theory and research (Mansbridge et al., 2010; Mendelberg et al., 2014). The negative association between facilitation and respect, however, stands out. It may be that when process organizers anticipate great difficulty with maintaining respect among participants, they employ facilitation. On the other hand, the dynamics of facilitation in deliberation have been understudied, with most attention on facilitation in mini-public designs. Generalization to other forms of participatory governance is risky and preliminary research suggests a wide variance in facilitator behaviours, which could account for this effect in the cases we studied. (See, for example, Dillard, 2013.) As we outline above this finding was not corroborated by all models, so requires further investigation.

Turning to the relationship between process and outcomes, we found that changes in participants' civic engagement were significantly associated with the quality of participants' collective decisions but also with the breadth of their deliberations, the latter robust across models. This accords with theories in which experiences of strong deliberation have the potential to transform citizen-participants (Chambers, 1996; Gutmann & Thompson, 1996, 2004). Moreover, the breadth of deliberations was positively and significantly associated with participants' satisfaction with their decision. Since such decisional satisfaction has been linked to deliberative participants' personal transformation (Gastil et al., 2010), the possibility that such satisfaction mediates the association between deliberative breadth and civic-engagement gains should be explored in future research.

In addition, it seems relevant here that respectful deliberation was linked to the quality of the decisions these bodies reached. These findings are also consistent with theories of “public judgment,” in which thoughtfully and respectfully “working through” the intricacies of policy issues, social conflict, and personal value commitments—a dynamic that is fostered by mutual respect among participants (Mathews, 1999, pp. 115, 130; Yankelovich, 1991, p. 165)—can bolster citizens’ sense of democratic civic identity.

On the path to policy influence, the consideration of different points of view was positively associated with a process’s decisions subsequently influencing public policy. This is consistent with accounts of deliberation that emphasize that its value to policymakers lies in its capacity to integrate a greater diversity of perspectives (Bohman, 2006; Fearon, 1998; Landemore, 2013). Further, depth of deliberative discussion was negatively associated with influence on policy, consistent with previous findings that policy makers often show little interest in the conclusions of citizens’ substantive policy deliberations (e.g., Smith, Richards, & Gastil, 2015).¹⁵

These findings are preliminary and require not only replication but also further model refinement and extension. The method employed in this study—the creation of a series of stand-alone linear regression models, which was necessitated by the small sample size and the low number of valid cases per item—does not permit the estimation of indirect and combined direct and indirect associations among variables. In future research, these issues may be addressed by analyzing data from a larger sample of Participedia cases by means of structural equation modelling (Druckman & Nelson, 2003; Scheufele et al., 2006). Nevertheless, our analysis here moves the research field over stubborn barriers to comparison of a greater diversity of cases.

Other limitations characterize the research reported in this article. The first limitation, noted above, concerns the validity of the data employed in the study: crowd-sourced case studies. Nonetheless, crowd-sourcing per se is no bar to obtaining valid data, since crowd-sourced information has been found to achieve acceptable validity in many domains (e.g., Heinzelman & Meier, 2013). Further, although case-study authors are subject to bias, the effect of such biases on case-study content can be diminished by systematizing the case-writing process and triangulating case details by requiring citations to multiple sources (Yin, 2014), techniques that are built into Participedia’s case-composition procedures.

Our study is also limited by other aspects of the data. These include contributors’ self-selection of cases documented in Participedia, the likelihood that cultural and contextual variations across the deliberations may yield quite different relationships among the variables,¹⁶ and the presence of outliers, non-normal frequency distributions, and other forms of “noise” in the data that required substantial cleaning procedures. We believe that many of these problems are likely to diminish as the number and variety of Participedia cases and case contributors increase.

Another limitation concerns low rates of intercoder agreement. Since inter-rater reliability is valued in part as a gauge of validity (Neuendorf, 2002), we have furnished an alternative measure of reliability that is recognized as a means of assessing validity: scale reliability (Rosenthal & Rosnow, 2008). We have also offered substantive (Haskard et al., 2009) and statistical (Good & Mittal, 1987; Samuels, 1993) explanations for the coincidence of low inter-rater reliability and adequate scale reliability.

Conclusion

This article is a response to a weakness in the literature on democratic innovations. Too often empirical analysis focuses on single case studies. When analysts move beyond this focus, the tendency has been to develop cross-case analysis within type (e.g., Deliberative Polls or participatory budgeting) or to engage in abstract cross-type comparative analysis of the implications of different design characteristics. The weakness of many of these studies is that they tend to focus on a limited pool of well-regarded processes. The development of Participedia represents a step-change for the analysis of democratic innovations, offering the opportunity to engage in more substantial cross-case and cross-type analysis in order to test claims within democratic theory. This paper is a first attempt to exploit meaningfully the diverse case material crowd-sourced on Participedia. In so doing it provides evidence not only that such analysis is possible, but also that the crowd-sourced data can be utilized to test assumptions in deliberative-democratic design.

As Participedia continues to develop, its capacity to support robust social science will be enhanced. The ongoing revision of the data-model used for case-submissions as well as introducing surveys for participants and observers of processes promises increasingly robust data over time. The organizers of crowd-source platforms like Participedia should pay careful attention to the comprehensiveness of the cases that appear therein. Insufficiently detailed cases pose problems such as those encountered in this study. Measurement of key variables can prove difficult when gleaned from vague case descriptions, and missing data problems can cause cases to drop from analysis altogether.

Moreover, researchers will always have to be mindful of Participedia's inherent biases of crowd-based platforms and adjust their analytic strategies accordingly (e.g., by applying suitable sampling matrices). That said, crowd-sourced data has contributed handsomely in other fields where cases cannot be reached in a timely enough fashion using traditional means (e.g. Bittner et al., 2016; Heinzelman & Meier, 2013). Participedia offers the opportunity to contribute to our understanding of contemporary democracy in similar ways. When we compare selected Participedia cases, we can test the predictive value of generalizations by asking whether they help diagnose important associations within cases on the platform, as well as cases that will be added in the future or are held by other researchers. As we get closer to a sampling frame for less problematic populations, at least within types of democratic innovations—as we may be approaching for example for mini-publics (Ryan & Smith, 2014)—probability sampling becomes a possibility. The expansion of the Participedia project will support analysis and insights that move us far beyond the current staple of single case studies and small-N comparisons.

Using an input-process-output model of democratic deliberation, we have been able to demonstrate the importance of design characteristics, in particular the important function played by information resources to enable deliberation. Our analysis of associations between variables suggests that there are distinct links between democratic respect, democratic consideration, and deliberative breadth, on the one hand, and outcomes such as decision quality, changes in participants' civic engagement, participants' decision ratings, and policy impact, on the other.

Crowd-sourced data on the Participedia platform have pointed to potential new lines of inquiry on deliberative design. These insights will be of interest both to practitioners engaged in building

participatory institutions and for refinements in theories of deliberative democracy. We entreat our colleagues to challenge or confirm these findings through the generation of further high-quality case studies and comparative research enabled by Participedia.

References

- Baiocchi, G. (2005). *Militants and citizens: The politics of participatory budgeting in Porto Alegre*. Stanford: Stanford University Press.
- Barrett, G., Wyman, M., & Coelho, V. S. P. (2012). Assessing the policy impacts of deliberative civic engagement: Comparing engagement in the health policy processes of Brazil and Canada. In T. Nabatchi, J. Gastil, G. M. Weiksner, & M. Leighninger (Eds.), *Democracy in motion: Evaluating the practice and impact of deliberative civic engagement* (pp. 181–203). Oxford: Oxford University Press.
- Belsley, D., Kuh, E., & Welsch, R. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley.
- Bigham, J., Bernstein, M., & Adar, E. (2015). Human-computer interaction and collective intelligence. In T. Malone & M. Bernstein (Eds.), *Handbook of collective intelligence* (pp. 57–84). Cambridge, MA: MIT Press.
- Bittner, C., Michel, B., & Turk, C. (2016). Turning the spotlight on the crowd: Examining the participatory ethics and practices of crisis mapping. *ACME: An International E-Journal for Critical Geographies*, 15(1), 207–229. Retrieved from <https://www.acme-journal.org/index.php/acme/article/view/1238>
- Bohman, J. (1998). Survey article: The coming of age of deliberative democracy. *Journal of Political Philosophy*, 6, 400–425. doi:10.1111/1467-9760.00061
- Bohman, J. (2006). Deliberative democracy and the epistemic benefits of diversity. *Episteme*, 3, 175–191.
- Burkhalter, S., Gastil, J. & Kelshaw, T. (2002). A conceptual definition and theoretical model of public deliberation in small face-to-face groups. *Communication Theory*, 12, 398–422. doi:10.1111/j.1468-2885.2002.tb00276.x
- Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Chambers, S. (1996). *Reasonable democracy: Jürgen Habermas and the politics of discourse*. Ithaca, NY: Cornell University Press.
- Chambers, S. (2003). Deliberative democratic theory. *Annual Review of Political Science*, 6, 307–326. doi:10.1146/annurev.polisci.6.121901.085538
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Delli Carpini, M., Lomax Cook, F., & Jacobs, L. (2004). Public deliberation, discursive participation, and citizen engagement: A review of the empirical literature. *Annual Review of Political Science*, 7, 315–344. doi:10.1146/annurev.polisci.7.121003.091630
- DeVellis, R. (2012). *Scale development: Theory and applications* (3rd ed.). Los Angeles, CA: Sage.

- Dillard, K. (2013). Envisioning the role of facilitation in public deliberation. *Journal of Applied Communication Research*, 41, 217–235. doi:10.1080/00909882.2013.826813
- Druckman, J., & Nelson, K. (2003). Framing and deliberation: How citizens' conversations limit elite influence. *American Journal of Political Science*, 47, 729–745.
- Fearon, J. (1998). Deliberation as discussion. In J. Elster (Ed.), *Deliberative democracy: Essays on reason and politics* (pp. 44–68). Cambridge, UK: Cambridge University Press.
- Fishkin, J. S. (2009). *When the people speak: Deliberative democracy and public consultation*. Oxford, UK: Oxford University Press.
- Font, J., Pasadas del Amo, S., & Smith, G. (2016). Tracing the impact of policy proposals from participatory processes: Methodological challenges and substantive lessons. *Journal of Public Deliberation*, 12(1), Article 3. Retrieved from <http://www.publicdeliberation.net/jpd/vol12/iss1/art3/>
- Fox, J. (2008). *Applied regression analysis and generalized linear models* (2nd ed.). Los Angeles, CA: Sage.
- Fung, A. (2003). Survey article: Recipes for public spheres: Eight institutional design choices and their consequences. *Journal of Political Philosophy*, 11, 338–367. doi:10.1111/1467-9760.00181
- Gastil, J. (2008). *Political communication and deliberation*. Thousand Oaks, CA: Sage.
- Gastil, J., Deess, E. P., Weiser, P. J., & Simmons, C. (2010). *The jury and democracy: How jury deliberation promotes civic engagement and political participation*. Oxford: Oxford University Press.
- Gastil, J., Knobloch, K., & Kelly, M. (2012). Evaluating deliberative public events and projects. In T. Nabatchi, J. Gastil, G. M. Weiksner, & M. Leighninger (Eds.), *Democracy in motion: Evaluating the practice and impact of deliberative civic engagement* (pp. 205–230). Oxford, UK: Oxford University Press.
- Gastil, J., Knobloch, K., & Richards, R. (2015). *Empowering voters through better information: Analysis of the Citizens' Initiative Review, 2010-2014*. Retrieved from <http://sites.psu.edu/citizensinitiativereview/wp-content/uploads/sites/23162/2015/05/CIR-2010-2014-Full-Report.pdf>
- Gastil, J., Rosenzweig, E., Knobloch, K. R., & Brinker, D. (2016). Does the public want mini-publics? Voter responses to the Citizens' Initiative Review. *Communication and the Public*, 1, 174–192. doi:10.1177/2057047316648329
- Good, I., & Mittal, Y. (1987). The amalgamation and geometry of two-by-two contingency tables. *Annals of Statistics*, 15, 694–711. doi:10.1214/aos/1176350369
- Graham, J. (2012). *Missing data: Analysis and design*. New York: Springer.
- Gutmann, A., & Thompson, D. (1996). *Democracy and disagreement*. Cambridge, MA: Belknap Press of Harvard University Press.
- Gutmann, A., & Thompson, D. (2004). *Why deliberative democracy?* Princeton, NJ: Princeton University Press.

- Haskard, K., DiMatteo, M., & Heritage, J. (2009). Affective and instrumental communication in primary care interactions: Predicting the satisfaction of nursing staff and patients. *Health Communication, 24*(1), 21–32. doi:10.1080/10410230802606968
- Heinzelman, J., & Meier, P. (2013). Crowdsourcing for human rights monitoring: Challenges and opportunities for information collection and verification. In J. Lannon & E. Halpin (Eds.), *Human rights and information communication technologies: Trends and consequences of use* (pp. 123–138). Hershey, PA: Information Science Reference.
- Kahane, D., Loftson, K., Herriman, J., & Hardy, M. (2013). Stakeholder and citizen roles in public deliberation. *Journal of Public Deliberation, 9*(2), Article 2. Retrieved from <http://www.publicdeliberation.net/jpd/vol9/iss2/art2/>
- Landemore, H. (2013). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton, NJ: Princeton University Press.
- List, C., Luskin, R. C., Fishkin, J. S., & McLean, I. (2013). Deliberation, single-peakedness, and the possibility of meaningful democracy: Evidence from Deliberative Polls. *Journal of Politics, 75*(1), 80–95. doi:10.1017/S0022381612000886
- Mansbridge, J., Bohman, J., Chambers, S., Estlund, D., Føllesdal, A., Fung, A., . . . Martí, J. L. (2010). The place of self-interest and the role of power in deliberative democracy. *The Journal of Political Philosophy, 18*(1), 64–100. doi:10.1111/j.1467-9760.2009.00344.x
- Mathews, D. (1999). *Politics for people: Finding a responsible public voice* (2nd ed.). Urbana: University of Illinois Press.
- Melville, K., Willingham, T., & Dedrick, J. (2005). National Issues Forums: A network of communities promoting public deliberation. In J. Gastil & P. Levine (Eds.), *The deliberative democracy handbook: Strategies for effective civic engagement in the twenty-first century* (pp. 35–58). San Francisco, CA: Jossey-Bass.
- Mendelberg, T., Karpowitz, C., & Oliphant, J.B. (2014). Gender inequality in deliberation: Unpacking the black box of interaction. *Perspectives on Politics, 12*(1), 18–44. doi:10.1017/S1537592713003691
- Mill, J. S. (1950). *Philosophy of scientific method*. (E. Nagel, Ed.). New York: Hafner.
- Mooney, C. Z. (1996). Bootstrap statistical inference: Examples and evaluations for political science. *American Journal of Political Science, 40*, 570–602.
- Nabatchi, T., Gastil, J., Weiksner, G. M., & Leighninger, M. (Eds.). (2012). *Democracy in motion: Evaluating the practice and impact of deliberative civic engagement*. Oxford, UK: Oxford University Press.
- Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.
- Participedia.net. (2013). *Participedia.net case data codebook*. [S.l.]: Participedia.net. Retrieved from http://participedia.net/en/export/csv/browse/cases_page/case?destination=browse/cases
- Pavitt, C. (1999). Theorizing about the group communication-leadership relationship: Input-process-output and functional models. In L. Frey (Ed.), *The handbook of group communication theory and research* (pp. 313–334). Thousand Oaks, CA: Sage.

- Pincock, H. (2012). Does deliberation make better citizens? In T. Nabatchi, J. Gastil, G. M. Weiksner, & M. Leighninger (Eds.), *Democracy in motion: Evaluating the practice and impact of deliberative civic engagement* (pp. 135–162). Oxford, UK: Oxford University Press.
- Polletta, F. (2015). Public deliberation and political contention. In C. W. Lee, M. McQuarrie, & E. T. Walker (Eds.), *Democratizing inequalities: Dilemmas of the new public participation* (pp. 222–243). New York: New York University Press.
- Richards, R., & Gastil, J. (2014). *Participedia case codebook*. University Park: Pennsylvania State University.
- Rosenberg, S. W. (Ed.) (2007). *Deliberation, participation and democracy: Can the people govern?* Houndmills, UK: Palgrave Macmillan.
- Rosenthal, R., & Rosnow, R. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). Boston, MA: McGraw-Hill.
- Ryan, M., & Smith, G. (2014). Defining mini-publics. In K. Grönlund, A. Bächtiger, & M. Setälä (Eds.), *Deliberative mini-publics: Involving citizens in the democratic process* (pp. 9–26). Colchester, UK: ECPR Press.
- Samuels, M. (1993). Simpson's paradox and related phenomena. *Journal of the American Statistical Association*, 88(412), 81–88. doi:10.1080/01621459.1993.10594297
- Scheufele, D., Hardy, B. W., Brossard, D., Waismel-Manor, I. S., & Nisbet, E. (2006). Democracy based on difference: Examining the links between structural heterogeneity, heterogeneity of discussion networks, and democratic citizenship. *Journal of Communication*, 56, 728–53. doi:10.1111/j.1460-2466.2006.00317.x
- Sintomer, Y., Herzberg, C., Röcke, A., & Allegretti, G. (2012). Transnational models of citizen participation: The case of participatory budgeting. *Journal of Public Deliberation*, 8(2), Article 9. Retrieved from <http://www.publicdeliberation.net/jpd/vol8/iss2/art9/>
- Smith, G. (2009). *Democratic innovations: Designing institutions for citizen participation*. Cambridge, UK: Cambridge University Press.
- Smith, G., Richards, R., & Gastil, J. (2015). The potential of Participedia as a crowdsourcing tool for comparative analysis of democratic innovations. *Policy & Internet*, 7, 243–262. doi:10.1002/poi3.93
- Thompson, D. F. (2008). Deliberative democratic theory and empirical political science. *Annual Review of Political Science*, 11, 497–520. doi:10.1146/annurev.polisci.11.081306.070555
- Wampler, B. (2007). *Participatory budgeting in Brazil: Contestation, cooperation, and accountability*. University Park: Pennsylvania State University Press.
- Warren, M. E. (1993). Can participatory democracy produce better selves? Psychological dimensions of Habermas's discursive model of democracy. *Political Psychology*, 14, 209–234. doi:10.2307/3791409
- Warren, M. E. (2009). Governance-driven democratization. *Critical Policy Studies*, 3(1), 3–13. doi:10.1080/19460170903158040

Warren, M., & Pearse, H. (2008). *Designing deliberative democracy: The British Columbia Citizens' Assembly*. Cambridge, UK: Cambridge University Press.

Yankelovich, D. (1991). *Coming to public judgment: Making democracy work in a complex world*. Syracuse, NY: Syracuse University Press.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Los Angeles, CA: SAGE.

Appendix

The Appendix is available online at: <http://www.publicdeliberation.net/jpd/vol13/iss2/art1/>

Notes

¹ University of British Columbia political science professor Mark Warren is the principal investigator of the grant, “Participedia: A Global Partnership to Create and Mobilize Knowledge about Democratic Innovations,” which runs until June 2020.

² Five cases in the sample—IDs 28, 37, 42, 140, and 456—employed methods of decision making (viz., taking the sense of the room or voting) along with participant opinion surveys.

³ The online appendix is available at: <http://www.publicdeliberation.net/jpd/vol13/iss2/art1/>

⁴ For example, the code measuring the clarity of the question put to the deliberating group read: “The charge or question was clearly put to the panelists,” with response options consisting of a five-point Likert-type scale having poles of “Strongly disagree” and “Strongly agree.” For more details on the coding process, see the online appendix.

⁵ Since uncertainty about the population of participatory-governance cases raises questions about the use of parametric inference, and because many statistical models fit to these data yielded low degrees of freedom, bootstrap regression was employed. As Mooney (1996, p. 570) writes, “Bootstrapping is a nonparametric approach to statistical inference” in which the sample is used as a population; an algorithm draws hundreds of probability samples with replacement from the initial sample and thereby produces a distribution of estimates which can be used for statistical inference. In this study, bootstrap regression was conducted in SPSS. The bootstrapping algorithm was configured for 1120 samples, with 95-percent confidence intervals estimated using the bias-corrected and accelerated method. In some samples, one or more variables were constant or had missing correlations, causing some estimates to be based on fewer than 1120 bootstrap samples. (See Tables 3, 4, and 5.) For results of *F* tests and estimated *Adjusted R*² for the OLS regression models on which bootstrap regressions were based, see Tables 15.1A and 15.2A in the online appendix.

⁶ As Belsley et al. (1980, p. 6; Fox, 2008, p. 241) write, outliers raise concerns because they “can have a disproportionate influence on the estimated parameters” and because “it is quite possible that the model-estimates are based primarily on” the outliers “rather than on the majority of the data.” Belsley et al. (1980) and Fox (2008) recommend multiple methods for detecting outliers. Of those methods, we applied several, on their own or in combinations, in order to identify outliers in our data, as detailed in “General Note for Tables 3A through 15.2A” in the appendix. These different methods often resulted in different observations’ being identified as outliers. We report results reflecting all of these methods, either in tables 3, 4, and 5 below or in Tables 3A through 15.2A in the appendix. For all but three of the models reported in tables 3, 4, and 5 below, “outlier residuals” were defined as those where (a) Cook’s *D* exceeded the recommended cutoff and the *t*-test for the Studentized residual had two-tailed *p* < .05, or (b) COVRATIO was below 1.0 and exceeded the recommended cutoff (see Belsley et al., 1980; Fox, 2008). For the other three models (i.e., the models of *Decision Quality*, *Participant Decision Rating*, and *Change in Participants’ Civic Engagement [Short-Term]*), “outlier residuals” were defined as those for which Cook’s *D* exceeded the cutoff stated in Belsley et al. (1980), or COVRATIO exceeded the cutoff recommended in Belsley et al. (1980). For the latter three models, the “Model G” combination of outlier-detection methods was used because it yielded clearer results than did the models featuring other outlier-detection methods, results of which appear in Tables 11A, 12A, and 14A in the appendix. In light of Fox’s (2008, p. 260) recommendation that “[a]lthough problematic data should not be ignored, they also should not be deleted automatically and without reflection,” we present alternative models from which no outliers were omitted (see appendix), and, regarding models from which we omitted outliers, outliers were first identified using multiple techniques—just described—recommended by Belsley et al. (1980) and Fox (2008) and then carefully considered before omission.

⁷ In cases such as these, where results have theoretical coherence but fall on the wrong side of the .05 criterion, we report results with *p* value < .10, along with coefficients and standard errors. This is consistent with recent scholarship that eschews “a fixed, dichotomous decision-level approach” to statistical significance (Rosenthal & Rosnow, 2008, p. 356).

⁸ This and all subsequent reported *p*-values reflect two-tailed tests.

⁹ Bootstrap regression yielded somewhat ambiguous inference results for the association between *Democratic Consideration* and *Deliberative Depth*. This association reached the conventional significance threshold (*b* = .57, *SE* = .26, *p* = .03, two-tailed) and confidence intervals estimated with the percentile method excluded zero (95% CI = [.03, 1.15]). Yet confidence intervals estimated with the bias-corrected and accelerated method (bias = .04) included zero (95% CI = [-.37, 1.50]).

¹⁰ As Fox (2008, p. 54) notes, many statistical procedures, including ordinary least squares regression, compare means of variables, but “[t]he mean of a skewed distribution is not ... a good summary of its center.” Fox (2008, p.

54) therefore recommends transforming each skewed variable in order to “make [the] skewed distribution more symmetric” and thus a better “summary of [the variable’s] center.” Accordingly, in this study, in some alternative models, variables with non-normal frequency distributions were transformed to render those distributions closer to normal. See skewness and kurtosis statistics for variables in Tables 1A and 2A and the section called “Additional Information About Interval and Index Variables” in appendix.

¹¹ In this section, “full corroboration” means, for a coefficient in Table 3, 4, or 5 that approached significance ($p < .10$), that that coefficient approached or reached significance ($p < .10$ or $p < .05$) in one or more alternative models, and for a coefficient in Table 3, 4, or 5 that reached significance ($p < .05$), that the coefficient reached significance ($p < .05$) in one or more alternative models; and “partial corroboration” means, for a coefficient in Table 3, 4, or 5 that reached significance ($p < .05$), that the coefficient approached significance ($p < .10$) in one or more alternative models.

¹² One association that was not corroborated at all by results of alternative models was the negative and significant association between *Deliberative Process Clarity* and *Democratic Consideration*. See Table 7A in appendix.

¹³ Moreover, the positive and significant association between *Stakeholders’ Role in Design Process* and *Materials Quality* (Table 3) was partially corroborated in two alternative models (Table 3A in the online appendix).

¹⁴ As Kahane et al. (2013, p. 21) state, stakeholders at times advocate reducing the numbers of expert or other neutral witnesses who are allowed to testify at deliberative events “in the name of giving stronger expression to the popular will or the interests of certain stakeholders.” Gastil et al. (2015, p. 11, note 14) explain that “[i]dentifying, recruiting, and selecting these experts is a significant expense for program staff, and pro and con advocates sometimes question the neutrality or relevance of such witnesses.” In response to those challenges, organizers of Citizens’ Initiative Reviews [CIRs] actually “eliminat[ed]... background witnesses” from CIRs held in the U.S. in 2014 (Gastil et al., 2015, p. 5).

¹⁵ As Barrett et al. (2012) note, public officials often use participatory processes instrumentally to place a veneer of popular legitimacy on policy decisions officials previously made without public input. In such circumstances, as Polletta (2015, p. 232) observes, officials are likely to embrace only a small subset of possible participatory-process outcomes, such as those that consist of “the policy that decision makers already wanted or” a policy “that [is] so unimportant as to not count as a concession of any kind.”

¹⁶ We thank an anonymous reviewer for contributing this point.