Supplemental Material

How Deliberative Designs Empower Citizens' Voices:

A Case Study on Ghana's Deliberative Poll on Agriculture and Environment

Kaiping Chen

I. Operationalizing speech acts

II. Computational analysis

III. Operationalizing deliberative reasoning from using and enhancing Discourse Quality Index (DQI)

IV. Inter-coder agreement

V. Details on different types of justification used by participants during deliberation

VI. Pre and post opinion changes

VII. Comparing participants' arguments in deliberation vs expert messages in the information video

I. Operationalizing speech acts

Transcripts from the small group deliberation were organized by the translated agency like the screenshot below. When I operationalized a speech act, each red box means one speech act. The character "D" refers to a deliberator (i.e., participant). In the screenshot below, there are three speech acts from the participants.

D: That would have been wonderful and would have helped us a lot. As I sit here, I am a farmer and I belong to a farmers group of over twenty two people. The problems we have with agriculture is that the people who are educated don't work on farms but rather want to sit in offices. They should rather have farms so that we those who have no education can learn firsthand from them. But this isn't the case. What is the use of being educated in agriculture and not farming? The situation is now such that if you are into farming then you aren't enlightened. Isn't this sad? This is what is making our food security issue a challenge. So for me, the agric officers should come out in their numbers so we can learn from them and we all benefit as a people.

Speech act 1

D: If the assembly can send out officers to guide farmers as to farming times, planting times and issues concerning their farms, that would help us a lot. I think it is a very good initiative if pursued.

Speech act 2

D: They have said it already. How can you have knowledge about farming and agriculture and want to go and sit in an office without farming? We the poor ones, we want to farm but have no capital and again have very little know how. The people in the offices have the capital and can go buy the fertilizers, the seeds, the weedicides and all. We the ones who really want to farm have no access to these things. How can we develop?

Speech act 3

II. Computational analysis

In this paper, I applied large-scale hand-coding and the supervised machine learning method to code all speech acts from the Tamale Deliberative Poll (DP) transcripts including 15 small group discussions taking place over two days.

The procedures are: 1.) two randomly chosen small group discussions from the Tamale DP were hand coded (n=422 speech acts, accounting for 13.69% of the total number of speech acts in Tamale DP). The number of speech acts to code for the training dataset was decided by ensuring that every variable in our coding scheme had samples for computational learning; 2.) to strengthen the validity of the findings from the manual coding, I employed several automated text analysis methods to computationally code the rest of the small group discussion transcripts from the Tamale DP. Automatic text analysis is a method that uses the assistance of computer programming to help researchers analyze a large-scale of text data such as transcripts, social media posts and newspaper articles. Scholars have developed a variety of methods for performing automatic text analysis (for detailed discussion, see Fig.1 in Grimmer and Stewart, 2013) including supervised, unsupervised machine learning methods, or a mix of two. In this paper, I used supervised machine learning (SML), a method for when researchers know what categories they want to include in the content analysis based on social science theories. I chose this method for this paper because I already knew which categories I wanted to use in the content analysis on the deliberation transcripts such as level of reasoning. My goal was to train the computer to learn from the hand-labelled dataset. The basic procedures of SML are: 1) a random sample of texts from all of the text data are labelled by humans following the categories developed by researchers (i.e. training dataset), 2) a machine learning algorithm is then selected to learn the features from this train dataset and during this process, cross-validation methods will be used to evaluate whether a classifier from the algorithm can predict the train dataset as accurately as human beings, 3) once the best classifier is decided, researchers will use it to code the rest of the dataset with the assistance of a computer programming software such as R or Python. In my case, I used the R software programming language. Researchers can install existing packages in R to perform different tasks. These packages are programming functions written by the researcher community to provide other scholars a ready to use code for a specific task. I used the *RTextTools* package (Jurka et al., 2013), which is designed for applying the supervised machine learning method to automatically code text data. This package provides a variety of machine learning algorithms such as supporting vector machine, boosting, bagging, random forest, and neutral network which differ in the ways these algorithms learn about the human-labelled data. We can think of these algorithms as our hired computer research assistants. As researchers, we show our computer research assistants how we labelled our train dataset, and then our computer research assistants will apply their unique learning strategies to code the rest of the dataset for us. To decide which computer research assistant is the best learner, I used the cross-validation function in the *RTextTools* package to perform the 10-fold cross validation to select a relatively better performed algorithm, boosting, which has an average accuracy[1] of 0.8 for my two levels of reasoning variables. I then applied the boosting algorithm to automatically code all of the rest speech acts. Results from the computational analysis of speech acts were similar to those from the manual content analysis. Supervised learning models are used in text mining to decrease the intensive labor required by manual content analysis (Barbera et al., 2020) and to improve consistency in content analysis.

Table 1

Tamale DP - Level of Reasoning in Opinions and Responses

(applied to **all** the 15 small group discussions for two deliberation days)

Opinions	No reasoning	7.6%
	Use reasoning	92.4%
Responses	Simple response	4.9%
	Substantive response	95.1%

III. Operationalizing deliberative reasoning using and enhancing Discourse Quality Index (DQI)

This paper drew from the “level of justification” in the DQI because this variable can help me analyze to what extent participants in the Tamale DP utilized reasoning. I operationalized the measurement of reasoning in two ways that differ from the “level of justification” in the DQI.

First, in the original DQI, level of justification has six values: *1) the speaker does not present any arguments (asks, for example, merely for additional information), 2) the speaker only says that X*

[1] Note: I also examined other performance evaluation metrics such as precision and recall for the level of reasoning variables. The precision is very high while the recall is low. The reason is because too few participants did not use reasoning when they expressed and responded to others in the train dataset. However, if we really want to obtain a train sample that can increase the recall performance, we need to hand label almost all speech acts and even if we hand label the full dataset, it is very likely that we still have too few observation on “not reasoning” for the computer algorithms to learn well. The nature of reasoning in the Tamale DP made it hard to use “recall” as a performance evaluation tool. However, for other datasets, I highly suggested examining all performance metrics and more importantly, which performance metrics to use depends on the goal of research. For instance, when there is a high cost associated with False Negative (e.g., sick patient detection), then recall is a very important criterion to consider; when the cost of false positive is high (e.g., spam detection), then precision is an important metric to consider. For detailed discussions, see: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

should or should not be done, that it is a wonderful or a terrible idea, etc. But no reason is given for why X should or should not be done, 3) the speaker justifies only with illustrations why X should nor should not be done, 4) The speaker gives a reason Y why X should or should not be done, but no linkage is made why Y will contribute to X, 5) The speaker gives a reason Y why X should or should not be done, and a linkage is made why Y will contribute to X, 6) The speaker gives at least two reasons why X should be done and for at least two reasons a linkage is made with X (see Appendix in Steiner, 2012). I simplified the coding on level of justification to two levels: whether a speaker uses justification or not because 1) I was interested in whether justification is used rather than the linkage between one's reasoning and one's preference and 2) the deliberation transcripts were translated from the Tamale local language to English: concerning the linkage part might get lost during translation or the way people link justification to argument might differ under various cultures, it is not suitable to study the linkage under this case.

Secondly, DQI does not have a variable which allows us to know the types of speech act: for instance, whether a speech act is an opinion expression from a participant or a response to others' arguments. Therefore, even with the variable "level of justification", it is difficult to distinguish whether the level of justification is made when one expresses opinions or when one responds to others. Making this distinction is crucial because as stressed by theories of deliberative democracy, deliberation is a process of "communicative reasoning," and this entails individuals use of reasoning when they exchange opinions. Responding to peers with grounded reasoning is the essence of deliberation. To measure this communicative reasoning aspect, I developed and used the indicator of "substantive response" in this paper, which captures whether people utilize reasoning when a speech act type is a response to others' arguments.

The importance of applying substantive response as an indicator is that it complements the limitations of the "respect" and "level of justification" indicators in the standard DQI. In the standard DQI, the authors attempt to measure intersubjectivity through the indicator "respect." Yet, through a deeper examination of the components of the respect indicator, I found that it greatly differs from "substantive response". In the latest version of the standard DQI (Steiner 2012: 269), Steiner distilled respect into three components: foul language, respect (respectful language), and respect (listening). An example of foul language is: *"you are a liar;"* and an example of respectful language is: *"your argument is truly brilliant."* Neither the foul nor the respectful language tells us if the speaker responds to other's arguments in a substantive manner. For instance, a reply such as *"Yes, I agree with you"* is coded as respect under standard DQI. However, it is a simple reply, which does not allow us to gauge if the speaker weights the merits of the competing arguments. Another component in Steiner's respect indicator is "respect (listening)", which measures how well a speaker listens to arguments addressed to them, but it does not cover how the speaker responds to other arguments, whether in a simple or a thoughtful manner. Next, the indicator of "level of justification" only captures to what extent one uses reasoning when one speaks (i.e., expresses their opinions). As such, this indicator cannot distinguish whether one's use of reasoning is a response to other opinions or is one's own opinion statement. Therefore, it is necessary to have a separate indicator (i.e., "substantive response"), in addition to the level of justification, to evaluate the quality of reasoning when one responds to competing arguments.

Considering these indicators' limitations, by using the indicator of "substantive response," I can measure if a speech act is a response to others' arguments, and whether it is a simple response or a substantive response. Examples of a simple response are: *"I agree; Yeah; Well, yes."* Conversely, a substantive response is defined as a reply that adds additional information to the other arguments or offers reasons to justify why a speaker supports/opposes the other arguments. An example of a substantive response is: *"I want to comment on what one of the ladies said. I disagree with education on radio. They should rather pay attention on strengthening the work of the sanitary officers. When someone is penalized for not keeping the surrounding clean, the neighbours will be deterred from doing same."* This is a substantive because the speaker responded to the previous speaker's opinion on education people about sanitation on radio by pointing out that it is more important to strengthen the work of the sanitary officers since these officers can deter neighborhood from bad sanitation practices.

IV. Inter-coder agreement

A total of seventy-nine speech acts (19% of all the hand-labelled speech acts) were hand-coded by two researchers to perform the inter-coder agreement check on the Tamale transcripts. I then used the R programming software and its open source package *irr* (Gamer et al., 2012) to calculate the Krippendorff inter-coder agreement score.

Coding Variables (unit: speech act)	**kripp.alpha**
Related to the Proposal	1.00
Type of Speech Acts	0.88
Opinion – Level of Justification	0.86
Response - Level of Reasoning	0.78

V. Details on different types of justification used by participants during deliberation

I read a random sample of the hand-labelled speech acts that were coded as use reasoning and then used inductive coding to map out what type of evidence participants used as justification. I found that the types of justifications can be summarized into four types and sometimes a participant used multiple types as justification. This analysis is inductive and did not involve inter-coder check. Therefore, evidence from Figure 1 in the main manuscript should be taken as suggestive. I provided the examples of each justification type below.

1). Using fact
e.g.: "I don't think it's important. This is because, laws in Ghana works only for a while."
2). Testimony/storytelling
e.g.: "*Susu is very important. I did savings with a "susu" company till I got money to pay for my child fees from the secondary school. So I think it's important and very good for me."*
3). Raising assumptions and conditions (for a proposal to work)
e.g.: *"If the assembly teaches us how to do this backyard farming, they should also teach us how to use good or clean water to water the vegetables so that the crops will grow well and healthy. It is not good to water crops with dirty or untreated water but people are likely to use dirty water if they don't know the effect of that."*
4). Raising suggestions (for the current proposal)

e.g.: *"I think the assembly should setup groups in areas so that they can meet them and give them this information rather than try to use mobile phones."*

VI. Pre and post opinion changes

Proposals	Before Deliberation	Post Deliberation	Significance
1. Promote training for households and community groups to set up backyard poultry farms	8.1	8.9	0.00
2. Promote training for households and community groups to set up backyard vegetable gardens	7.1	8.3	0.00
3. Promote access to information on credit opportunities for livelihood activities	7.6	8.3	0.00
4. Promote the setting up of village savings and loans associations	7.8	8.0	0.39
5. Provide water tanks for setting up rain water harvesting systems in all educational institutions	8.9	9.4	0.00
6. Promote access to credit for urban farmers through the Common Fund	7.9	8.5	0.01
7. Provide timely weather forecasting information for farming	8.6	8.8	0.11
8. Provide timely extension services for farming	8.8	9.0	0.09
9. Provide appropriate training for food storage	8.7	9.2	0.00
10. Provide technology training for food storage	8.6	9.3	0.00
11. Promote maximum use of local foods	8.8	9.3	0.00
12. Train people to prepare nutritious foods using local food items (millet, groundnuts)	8.6	9.10	0.00
13. Promote the cultivation of fonio and other neglected nutritious local crops	8.1	9.0	0.00
14. Promote food fairs to encourage the consumption of local foods	8.9	8.7	0.00
15. Promote the setting up of a mobile phone platform for providing information to farmers	7.7	7.7	0.75
16. Set up sewage treatment plants for managing solid and liquid waste	8.7	9.2	0.00
17. Encourage a Public-Private-Partnership to convert waste to energy	8.6	9.2	0.00
18. Ban the use of plastic carrier bags in the city	6.8	7.6	0.00
19. Promote the use of carrier bags made of biodegradable materials	8.1	9.0	0.00
20. Encourage media houses to allocate weekly airtime for water, hygiene and sanitation information	8.9	9.2	0.04
21. Promote the segregation of household waste by providing waste bins	8.5	8.7	0.33
22. Promote the sorting of waste by all institutions	8.6	8.9	0.09
23. Promote the use of environmentally-friendly toilets in all houses	9.3	9.5	0.05
24. Promote the use of environmentally-friendly toilets in all institutions	9.2	9.4	0.06
24b. Ban the setting up of vegetable farms within 100m of toilet facilities	8.6	8.8	0.36
25. Ban the use of untreated waste water for vegetable farming	8.5	9.1	0.00
26. Intensify the behavior change communication campaign to improve hygiene and sanitation	8.9	9.2	0.02
27. Intensify the hand washing campaign in schools	9.3	9.5	0.01
28. Build the capacity of local institutions such as the School of Hygiene to promote good hygiene and sanitation practices	9.0	9.4	0.00

Proposals	Before Deliberation	Post Deliberation	Significance
29. Promote a low cost treatment of waste water for farming through the use of charcoal and stones	7.8	8.4	0.00
30. Promote the use of drip irrigation	8.4	9.0	0.00
31. Encourage communities to use organic materials in agriculture such as composting	8.8	9.4	0.00
32. Promote the setting up of irrigation facilities adapted for urban settings such as using boreholes, wells and dugouts	8.6	8.9	0.07
33. Provide water tanks for setting up rain water harvesting systems in residential facilities	8.8	9.2	0.01
34. Ensure regular desilting of gutters	9.2	9.5	0.01
35. Construct and maintain gutters	9.2	9.3	0.38
36. Provide more opportunities for the most vulnerable to buy insecticide treated bed nets at a low price	8.9	9.1	0.10
37. Implement a systematic plan to control mosquitoes	9.2	9.5	0.00
38. Provide the most vulnerable with treated bed nets at a low price	8.9	9.1	0.11
39. Promote public education for effective cholera control	9.5	9.7	0.00
40. Some people think that vegetable farms should produce as much as possible, even if they have to use the waste water from the toilets (at point 0). Other people think that vegetables should only be produced with clean water, even if that means that many fewer vegetables are produced (at point 10).	9.0	9.5	0.00
41. How strongly would you agree or disagree with the following statements? 0 is strongly disagree, 10 is strongly agree and 5 is exactly in the middle			
41a. Providing water tanks for setting up rain water harvesting systems would ensure availability of more good quality water	8.6	9.0	0.01
41b. Treating waste water for farming would allow people to use good quality water for drinking	7.8	8.8	0.00
41c. Sewage treatment plants would prevent fecal and solid waste ending up in the wrong places	8.6	9.0	0.01
41d. Teaching people how to prepare nutritious meals would solve food contamination	8.1	8.9	0.00
41e. The media promotions on sanitation and hygiene will not change people's behavior	4.2	4.4	0.57
41f. Getting more access to loans and credit will improve our daily lives	7.2	7.6	0.05

Note: all questions are measured from a 0 to 10 scale, where 0 means extremely unimportant and 10 means extremely important.
Source: this table is obtained from Appendix Table C from (Chirawurah et al., 2019). I adjusted the decimal points of the scale on public opinions changes to round to one digit to meet the requirement from Public Understanding of Science.

VII. Comparing participants' arguments in deliberation vs expert messages in the information video

Expert Core Messages and Participants' Arguments

Issue	Expert core messages in the stimulus video	Main arguments raised by participants during deliberation
Livelihood & Food Security	1.Backyard farming is good	1.It is important to receive education about how to keep and care for animals 2.Backyard farming can help us sell farm produce to support our children 3.It is important for assembly to teach us how to use clean water to water our vegetables
	2.Nutrition is an important aspect to feed children. Mum should be taught on how to grow local foods	1.It is important to teach moms how to prepare nutritious food
	3.Empower mother with financial resources	1.There is a worry about the assembly providing loans because people might run away with these money 2.However, participants suggested that setting up the center (susu) for saving could be beneficial because susu is from people's own saving 3.Assembly should put their words into actions

References

Barberá, P., Boydstun, A. E., Linn, S., McMahon, R., & Nagler, J. (2020). Automated text classification of news articles: A practical guide. *Political Analysis*, 1-24.

Chirawurah, D., Fishkin, J., Santuah, N., Siu, A., Bawah, A., Kranjac-Berisavljevic, G., & Giles, K. (2019). Deliberation for Development: Ghana's First Deliberative Poll. *Journal of Public Deliberation*, *15*(1).

Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Package 'irr'. *Various coefficients of interrater reliability and agreement*.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, *21*(3), 267-297.

Jurka, T. P., Collingwood, L., Boydstun, A. E., Grossman, E., & van Atteveldt, W. (2013). RTextTools: A Supervised Learning Package for Text Classification. *R Journal*, *5*(1).

Steiner, J. (2012). *The foundations of deliberative democracy: Empirical research and normative implications*. Cambridge University Press.