Increasing Inclusion and Time-Efficiency in Participatory Policy-Making Deliberations with E-Scribing Technology

Gustavo Umbelino
gustavo@u.northwestern.edu
Northwestern University
Evanston, IL, USA

Kristine Lu
k.lu@u.northwestern.edu
Northwestern University
Evanston, IL, USA

Matthew Easterday
easterday@northwestern.edu
Northwestern University
Evanston, IL, USA

ABSTRACT

Citizens can increase openness, transparency, and accountability of institutions by taking part in face-to-face participatory policy-making deliberations, such as participatory budgeting assemblies. But for participants' contributions to influence policy outcomes, organizers need to capture and synthesize participants' input. Existing approaches are not inclusive for participants or require too much time from organizers. We designed *e-scribing*, a novel approach for capturing and synthesizing participants' input from face-to-face deliberations in real time by combining scribes with digital technology. To evaluate the approach, we built *DeliberationWorks*, a digital deliberation technology that helps scribes (a) capture proxy input (i.e., as participants) that is complete and accurate so that participants do not need to interact with technology themselves and (b) synthesize the discussion in real time using labels. We deployed DeliberationWorks with 5 scribes in two face-to-face deliberations with 8-10 participants and found that, on average, 82% of the input was captured mostly accurately. After one hour of training, scribes synthesized input within 10 minutes of the end of the deliberation. Our findings suggest that e-scribing makes participatory policy-making more inclusive by allowing participants to share their input without interacting with technology, and more time-efficient by reducing synthesis and training times for organizers.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; • **Applied computing** → *Computing in government*.

KEYWORDS

face-to-face, participatory policy-making, deliberation, scribe

ACM Reference Format:
Gustavo Umbelino, Kristine Lu, and Matthew Easterday. 2023. Increasing Inclusion and Time-Efficiency in Participatory Policy-Making Deliberations with E-Scribing Technology. In *16th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2023), September 26–29, 2023, Belo Horizonte, Brazil.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3614321.3614349

ICEGOV 2023, September 26–29, 2023, Belo Horizonte, Brazil

ACM ISBN 979-8-4007-0742-1/23/09…$15.00
https://doi.org/10.1145/3614321.3614349

1 INTRODUCTION

Citizens can increase openness, transparency, and accountability of institutions to address pressing social issues, like climate change and racial inequities, by taking part in face-to-face participatory policy-making deliberations—meetings where participants deliberate to improve or create policies to address problems in their community [1, 4, 12, 37–39].

For participants to influence policy outcomes, organizers must capture and synthesize participants' input to incorporate it into written policy proposals. Input must be captured in an inclusive manner to ensure that all participants are heard and that resulting policies reflect all participants' views [1, 26]. Input must also be synthesized in a time-effective manner to ensure that organizers can spend their time in other parts of the process, such as recruiting participants and training volunteers [15].

Unfortunately, existing approaches of capturing and synthesizing participant input in face-to-face contexts [e.g., 2, 21–23, 40, 42, 43, 45] can exclude participants or require a great deal of organizers' time. For example, the most common way to capture input is to have participants interact directly with digital technologies (e.g., Decidim, CONSUL DEMOCRACY[1]), which excludes those who are unwilling to or uncomfortable interacting with technology [29, 44]. Alternative approaches, such as artificial intelligence-based technologies or scribes, can help increase inclusion, but they are time consuming for organizers to synthesize the input captured during face-to-face deliberations.

To address this gap, we present *e-scribing*, a novel approach for capturing and synthesizing input from face-to-face participatory policy-making deliberations by having scribes capture structured input in real time using digital technology. To evaluate the e-scribing approach, we built *DeliberationWorks*, a digital deliberation technology designed to help participatory policy-making organizers to capture input from participants in an inclusive manner and to synthesize the input in minimal time by combining scribes and technology.

2 BACKGROUND

2.1 Participatory policy-making deliberations are increasingly popular

Participatory policy-making initiatives, such as participatory budgeting [38], participatory design [37], deliberative polls [12], and citizens assemblies [4], have spread around the world because they deepen public participation in decision-making [39] and improve the openness, transparency, and accountability of institutions [26].

[1] decidim.org, consulproject.org

Participatory policy-making often involves face-to-face deliberations where participants discuss and share ideas, reasons, suggestions, questions that organizers can use to improve or create new policies [29]. For example, in participatory budgeting, participants attend deliberations to generate ideas and decide how to spend part of a city's budget [e.g., 29]. After the deliberation, organizers synthesize participant input into reports to share with the community for transparency and with decision makers for implementation [18].

Despite recent advances on e-participation research and practice [36], face-to-face deliberations are a necessary part of participatory policy-making because they allow participants to build trust, commitment to the process, and overall have a higher quality deliberation [1, 5, 6, 20, 29, 34, 41]. Face-to-face deliberations also avoid the challenges of online technologies that can exclude populations [8, 29], especially marginalized groups such as older adults [44], or those who have difficulties interacting through digital technologies or written communication.

Unfortunately, it is still difficult for organizers to capture and synthesize input in a way that is inclusive for participants and does not take too much time [8, 30]. In this paper, we refer to inclusive input as input that is captured without participants having to interact with any digital technology, accurately captures the meaning of what a participant shared (i.e., accurate), and captures all inputs from each participant (i.e., complete). To synthesize the input quickly, input capture and synthesis should occur as simultaneously as possible, reducing the need to process input after the deliberation. To do so, organizers need to train scribes to segment and extract relevant input and label it in real time.

2.2 Existing approaches for capturing and synthesizing community input

Digital governance and human-computer interaction (HCI) researchers have proposed approaches to help organizers capture [22, 40, 43] and synthesize input from participants [2, 21, 23, 42, 45]. Existing digital technologies currently support three kinds of approaches for capturing and synthesizing input from face-to-face deliberations: participant-generated input, artificial intelligence-based, and scribe-based approaches. These approaches help increase participation in participatory policy-making initiatives, but do not capture input in both an inclusive [20, 29] or timely manner [8, 30] in face-to-face deliberations.

2.2.1 Participant-generated input approaches. The majority of existing technologies designed for capturing and synthesizing participant input generally rely on participants to provide the input themselves by interacting directly with technologies [8, 22, 23, 40, 42, 43, 45]. For example, digital technologies like *Wikium+* and *Tilda* allow participants to synthesize their own input while engaging in online discussion [42, 45] and could be employed in face-to-face contexts if participants have access and are willing to interact with digital devices. However, requiring interaction with digital devices excludes those who face challenges doing so [8, 44].

2.2.2 Artificial intelligence-based approaches. AI-based approaches that include natural-language processing (NLP) and machine learning (ML) help organizers collect all verbal input and minimize synthesis times [2, 10, 22]. Yet, in practice, these systems are not accurate or reliable enough to save time [2, 10, 24, 28, 32]. For example, automated speech recognition (ASR) technologies, such as Otter.ai[2], help organizers capture speech from face-to-face deliberations and automatically convert it to text, but organizers still spend a significant amount of time fixing inaccuracies in the transcriptions and synthesizing the input. Even when there are large amounts of accurate data available for training NLP models, it is unclear whether automated synthesis is consistent enough to save organizers' time [2].

2.2.3 Scribe-based approaches. Scribes can be trained to capture participant input at face-to-face deliberations [17, 19], but training scribes to capture specific types of input takes time. For example, volunteer scribes have been trained to capture input at large-scale, face-to-face deliberations while a team of volunteers synthesized their notes in real time [19], but training took a full day, which is a prohibitive amount of time for most volunteers.

Scribing technologies are too expensive and require too much training time [28]. For example, professional scribes who provide live-transcription for deaf and hard of hearing participants use special hardware that is too costly for capturing discussion notes in participatory policy-making deliberations [24, 28]. To address this challenge, prior research in crowdsourcing has employed crowdworkers to capture input in real time [28]. However, hiring crowdworkers is still costly due to the extra hardware (e.g., microphones) and the cost of hiring multiple workers. Moreover, full transcriptions are not necessary nor ideal for policy development, which benefits from more concise input structured as questions, reasons, suggestions, for example.

2.3 E-scribing: a novel approach for capturing and synthesizing input from participatory policy-making deliberations

The digital governance and the HCI communities want organizers to capture input from participatory policy-making deliberations in an inclusive manner for participants [8] and to synthesize the input in minimal time [2].

Existing theoretical and practical approaches, such as participant-generated input, AI-based technologies, and scribes alone are not sufficient to support inclusive participation at face-to-face deliberations without increasing the time needed from organizers.

To address this gap, this paper contributes *e-scribing* to the field of digital governance, a novel approach that combines scribes and scribe-support technology for capturing and synthesizing input from face-to-face participatory policy-making deliberations.

To evaluate the e-scribing approach, we designed and built *DeliberationWorks*, a digital deliberation technology with scribing features like *proxy commenting* and *comment labels* that help scribes capture structured input for participants without having them interact with technology while also labeling input in real time, thereby minimizing the time needed for training scribes to synthesizing input.

[2]otter.ai

3 DELIBERATIONWORKS SYSTEM DESIGN

Our goal in this project was to design a sociotechnical system (i.e., process and technology) for participatory policy-making that allows a representative group of participants to generate and vote on proposals without requiring extensive amounts of time or technological literacy.

Participants in previous participatory policy-making initiatives can take days, weeks, or months to generate and develop proposals for voting. As a result, only those with ample free time and resources can participate and reap benefits of that participation. To shorten the time required for policy-making, we split the work of decision making and policy development. A representative group of participants, the *decision makers*, attend two one-hour deliberations: an agenda-setting deliberation and a voting deliberation. In the agenda-setting deliberation, the decision makers generate and select a small number of ideas to develop into full proposals. In the voting deliberation, decision makers vote on which of the full proposals to implement. A second group of participants, the *policy developers*, with more time and expertise to develop the policy proposals that the decision makers selected for development. Splitting decision making from policy development can be thought of analogously to the division of labor between city council members and staff, and creates an inclusive policy-making process for a representative group of decision makers.

However, this split leads to two technical challenges: *technology exclusion* and *training time*. We overcome these challenges using the e-scribing approach, which combines scribes with digital technology to capture and synthesize participants' input in real time.

3.0.1 Technical challenge 1: Avoiding technology exclusion with scribes. Not every participant feels comfortable providing input in writing or online [44], and providing access to hardware is also prohibitively expensive. To overcome this challenge, e-scribing includes scribes to capture deliberation participants' input without requiring participants to directly interact with technology. To do so, scribes use *proxy commenting* to take notes as participants. Scribes use the scribing interface to create a group with the names of participants in their tables. Then, scribes write down input as participants deliberate (Figure 1, 1).

DeliberationWorks allows scribes to capture and synthesize input that is accurate and complete enough to influence policy development. The scribe interface was designed so that scribes could synthesize participant input with minimal effort. For example, with less than five clicks, scribes are able to label input captured with author, policy, and type labels. Making it easy to label input increases scribes' accuracy and completeness by allowing them to focus on the content of their notes and labels rather than on the process for capturing and labeling input.

3.0.2 Technical challenge 2: Minimizing training time. Scribes need to be trained in a brief amount of time to capture notes that are immediately useful for policy developers. To overcome this challenge, e-scribing includes digital technology designed for scribes and an *asynchronous training guide* for scribes to quickly learn how to synthesize input collected in real time. Although the exact time it takes to synthesize participants' input varies across contexts, it is widely agreed upon that this is one of the major challenges for organizers of participatory policy-making [21, 30].

Building on prior technologies that attempt to label participants' input in real time [22, 42, 45], we provide *comment labels* that scribes use to categorize the type of comment, the policy the comment is referring to, and the author of the comment. Comment labels allow scribes to quickly learn what input to capture and how to label it, saving organizers time to find and aggregate specific pieces of input. To label their notes, scribes click on the type of input using the type labels dropdown (Figure 1, 2), click on the author of the comment using the author labels (Figure 1, 3), and the policy label by clicking on the policy from the policy database (Figure 1, 4) or by typing in the policy number (Figure 1, 5).

The purpose of scribes' captured input is to support policy developers in developing policies off decision makers' ideas without having to be at the deliberations themselves. For example, when comments are labeled as *questions*, policy developers can easily spot those in the scribe notes and improve the policy by incorporating answers to participants' questions. Policy labels are used to link the comment to a policy from a policy database populated prior to discussion, which allow policy developers to filter input that was relevant to the policies they are responsible for developing. Author labels are used to attribute the author to the comment, which allow policy developers to contact participants directly in case they had clarifying questions about their input.

To reduce the time that it takes to train scribes to both capture and synthesize input in real time, we provided a self-paced asynchronous training guide, a slide presentation that scribes used to get comfortable with their task prior to the real deliberation. The training guide included an overview of the deliberation and DeliberationWorks, the codebook that scribes used to label participants' comments, and a sample deliberation video that they used to practice scribing. This training allowed scribes to quickly understand what input to capture the deliberation and how to do so using DeliberationWorks. Scribes were instructed to type in summarized notes into DeliberationWorks rather than verbatim transcripts to reduce the skill needed to type participant input and hence reduce the training time. Concise notes are desirable because they reduce the time of further synthesis of the deliberation.

4 METHOD

4.1 Context

We draw on data collected during the pilot stages of our city's first participatory budgeting process: a process that allows community members to decide how to spend part of the city's budget [38]. From Fall 2021 to Winter 2022, our research team ran two one-hour, face-to-face deliberations with undergraduate students as participants and scribes, and members of our team as facilitators. In the first deliberation, participants were given a set of 68 short policies compiled from various sources including canvassing with community members and city documents, and included policies such as *"provide microloans to small businesses"* and *"increase funding for mental health programs in schools."* These policies were categorized into 10 areas, such as *Education* and *Healthcare*. Moderators instructed participants to choose their top policy, provide their reasoning, ask questions, and suggest edits so that policy developers could

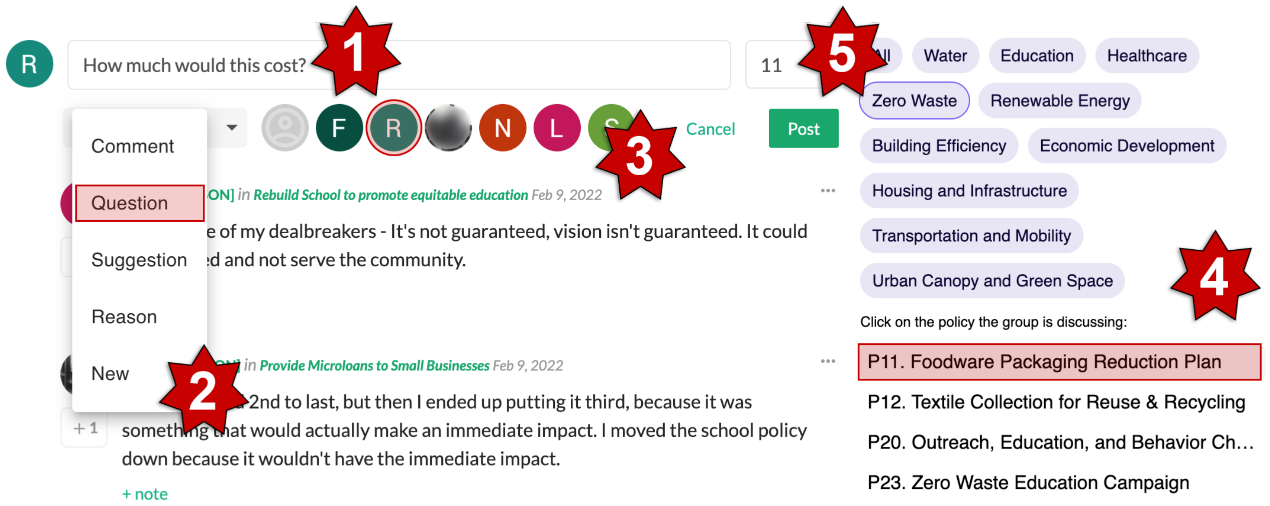

Figure 1: DeliberationWorks scribe interface allows scribes to capture input for participants using proxy commenting (1) and synthesize input in real time using comment labels (2-5).

develop the short policies into more complete proposals for the second deliberation. In the second deliberation, we invited the same participants back to individually rank 10 developed proposals and discuss their reasoning. During both deliberations, participants were divided into small groups of 4-8 people with one moderator and 1-2 scribes.

4.2 Participants

We recruited 16 undergraduate students at a Midwestern US university. Ten students were participants and six were trained as scribes. Eight out of the 10 participants came back for the second day of deliberation. Out of all students, seven were recruited from a community engagement class and the other 9 were recruited by the class students.

4.3 Data collection

We recorded audio recordings, surveys, and DeliberationWorks log data from the two deliberations and scribe trainings. Each deliberation took 60-70 minutes, and were audio recorded and transcribed with an automatic speech recognition (ASR) system for analysis. We also recorded survey responses from scribes after the training and after the deliberation. Two scribes responded to both surveys. In the post-training survey, scribes were asked how long the training took and how confident they were to scribe at a real deliberation. In the post-deliberation survey, scribes were asked how long it took them to finish editing their notes after the deliberation was over, and how long it took them to feel comfortable scribing before the deliberation.

4.4 Analysis

We analyzed completeness and accuracy of the scribe notes as compared to the audio transcript, similar to prior research on note-taking [7, 13, 14, 25]. We began the analysis with a researcher segmenting the audio transcripts into idea segments, or units [9], as if they were scribes for the group with access to the "ground truth" of what was said. We use segments as the unit of analysis rather than words, unlike prior research on live captioning scribes [28], because full verbatim transcripts are not necessary for proposal development in our context. The researcher labeled each segment with author, policy number, and type labels: *comment*, *question*, *suggestion*, *reason*, *new* following the codebook on Table 1. These labels were created prior to data collection and used to train the scribes.

To analyze *completeness*, we compared coded segments from the ground truth transcript with scribe notes and coded each segment as present or missing from the scribe notes. As a secondary analysis, we categorized the missing comments following a simple iterative coding process [33] to understand the kinds of input that scribes missed.

For *accuracy*, we compared the author, proposal number, type, and the content of each scribe note to the coded ground truth. Author, proposal, and type were analyzed as binary variables (i.e., accurate, inaccurate). For content, similar to prior note-taking research [e.g., 7, 14], one coder began by coding each scribe note as *mostly accurate*, *missing info*, and *inaccurate* (see Table 2). Then, two other coders did several rounds of coding on subsets of the data to improve the code descriptions. Once the codes were improved, two independent raters coded 42 of the 98 codes (43%) in the dataset and reached a substantial level of agreement indicated by the Cohen's Kappa statistic (K = 0.63) [27, 31]. To reduce observer bias (i.e., to avoid impartial judgements about the accuracy of scribe notes), one of the independent coders selected for this analysis was a trained qualitative researcher not involved in this project. Recognizing that most scribe notes were coded as mostly accurate (84%) and none as inaccurate (0%), we also calculated a prevalence and bias-adjusted Kappa statistic, which indicated an almost perfect level of agreement (PABAK = 0.86) [3, 27]. Seven notes were excluded from the accuracy analysis because they were too difficult for the researchers to understand and thus to compare to the scribe notes due to the

Table 1: Comment type labels used to train scribes and to code audio transcriptions

Type	Description	Example
question	clarifying questions about the proposal or the process	*It says increased funding for mental health programs, but there's already been adult programs that exist. Has there been a deficit?* (P5)
suggestion	suggestion for changing the proposal or the process	*I'd switch 'environment education' for 'critical race theory'* (P2)
reason	reason why participants support, oppose, or are unsure about a proposal	*I think it's important that there's education in the [district], but I think there's more pressing issues and the budget would barely put a dent in it* (P4)
new	new proposals participants come up with	*Rebuilding the school in the most underserved [district] would improve educational equity* (P3)
comment	comments that do not fit other categories	*I'm just surprised we don't have a 24-hour mental health crisis hotline* (P9)

noise in the environment. Five scribes were included in this analysis because one scribe did not participate in the deliberations despite being trained and filling out the post-training survey.

5 FINDINGS

5.1 Most input was captured and labeled without requiring participants to interact with technology

We found that in two one-hour deliberations, scribes captured on average 82% of all the input that participants shared out loud in discussion. In total, scribes captured 105 comments out of 128 shared. All scribe notes were labeled with the author of the comment, the policy it referred to (if any), and the type of comment, including 47 labeled as reasons, 16 suggestions, 9 questions, 5 new policies, and 24 general comments. None of the participants interacted with the technology to have their comments captured and labeled for further use by organizers.

The input that scribes failed to capture was generally input that was considered not useful for policy development and hence scribes were not trained to capture. For example, scribes missed questions that were answered by moderators or participants (8 questions), questions that happened before the discussion started (2 questions), comments contributing personal knowledge or reactions (e.g., *I think that [this policy] is really popular here within the [city's] community*; 7 comments), or indecisions about policies (3 comment). Only four of the comments missed should have been categorized according to the training guide, which suggests that it is possible that the majority of what was not captured was because scribes were following the training rather than due to distraction.

On the other hand, two of the comments captured by scribes were not captured by the automated transcription system because of the noisy environment and the placement of the external microphones relative to the speakers. The researchers also considered the the audio recording unintelligible.

5.2 The scribe notes captured were generally accurate, few missed important information

We found that, on average, scribes labeled 77% of their notes with the accurate type (e.g., question, reason, suggestion). There was a significant of variance between the scribes: the most accurate scribe labeled 100% their comments with the accurate types, while the least accurate scribe accurately labeled 50%. For example, when capturing P4's the reason for supporting a policy, scribe S3 wrote:

> *I worry this [policy #21] is inaccessible to some communities because it's more expensive. I think public transport is a better solution* (P4, #21, comment, scribed by S3)

In the above quote captured and labeled by S1 as comment, P4 explained their reasoning for not supporting policy #21, which is because of the cost of the policy. According to the coding book provided to the scribes during the training (see Table 1), this quote would have been more accurately labeled as a *reason* instead of *comment*. In terms of the author, scribes labeled 94% of their notes accurately, ranging from 100% to 90% accuracy. The high accuracy could be attributed to how participants introduced themselves at the beginning of the deliberation and scribes were instructed to add them to their groups in DeliberationWorks. Some of the participants also had profiles in the system and thus had profile pictures that scribes could use to identify their group members. Another factor contributed to the high accuracy was that some scribes already knew each other. Nevertheless, there were a few cases where scribes mislabeled the author even though they knew their names, suggesting that the scribe was distracted when taking their notes. One factor that might have compromised the accuracy of the author tags was that we did not provide nametags to participants, so scribes did not have a way to remember the names of participants in their groups unless they asked again. However, the groups were relatively small, ranging from 4 to 8 people, which made it relatively easy for the scribes to remember everyone's names.

In terms of policy number, scribes labeled 87% of their notes accurately, ranging from 100% to 74%. The lower accuracy, compared to author labels, might be explained by how participants often talked about multiple policies at the same time, making it unclear which policy they were commenting on. For example, when discussing their indecision about which policy to choose to support, P6 said:

> *I chose healthcare. I was torn between numbers #63, or Alternatives to the mental health emergency. But then also, like #65, Increasing access to mental health sources. Yeah, they are really important [unintelligible]. I also don't know a lot about [unintelligible], but I feel like homelessness is, I just think it's extremely important. I also think that [unintelligible] I feel like homelessness is [a problem] across America* (P6)

This quote was difficult to understand due to the noisy environment, but generally, the participant expressed indecision about policies #63 and #65, which are under the Healthcare category, as well as unmentioned policies under the Housing and Infrastructure category, which was mentioned by the previous participant in the discussion. Overall, P6 supports these policies because of how important these issues are nationally, but they feel they do not have enough knowledge about them. The scribe for the table, S5, captured:

> *It's extremely important. Although I don't know much about [the city], homelessness is a nation-wide issue, but we need people to care about mental health crises* (P1, #63, reason, scribed by S5)

S5 accurately labeled the type of note as *reason*, although the author was captured inaccurately (P1 instead of P6). S5 chose to label the comment with policy #63, one of the policies mentioned by the participant. This was not considered an accuracy mistake, but rather an example of inflexibility in the system or in the training.

In terms of content, we found that of the 98 scribe notes that were compared to the "ground truth" transcript, on average, 85% were coded as mostly accurate, 15% were missing information, and 0% were inaccurate. Scribe notes were coded as missing information rather than mostly accurate if they failed to capture concrete details that would help policy developers make policies more specific. For example, when speaking about the lack of a policy in Education to make high schools more equitable, P9 said:

> *I didn't find one that I believe serves what I believe to be the biggest need in [Education], that being equitable education across all [unintelligible] high schools for all the districts within [the city]* (P9)

The scribe failed to capture the specific context behind P9's reasoning, *high schools*:

> *It was difficult to pick a policy within education in [the city] that serves everyone in all districts/schools. It needs to be broad, equitable assistance and policy* (P9, comment, scribed by S5)

Unexpectedly, we found that scribes were able to capture *intent*, which would not be captured with automated transcription software. In one of the notes captured by S2, P3 had said that policy #21 would be a good if it was free:

> *I think each [policy] is the step to get to the next one. Yeah, so #21, like #15, is good if you make it free. Like, it's already much more affordable than having their own car* (P3)

Yet, as judged by the names and descriptions in the policy database, P3 really meant #22 instead of #21, which was captured by the scribe S2:

> *#22 is good if it was free. #15 is a good first step* (P3, #15, reason, scribed by S2)

5.3 DeliberationWorks helped organizers save time synthesizing participants notes and training scribes

We found that all the captured scribe notes were labeled within 10 minutes after the deliberation was over based on self-report from the scribes in the post-deliberation survey.

From the two scribes who responded to the post-training survey, we found that scribe training took at most one hour to complete; we also learned that both scribes took about two hours preparing for the deliberation, including the training they completed. This suggests that scribes took one hour after the training was over to get familiar with DeliberationWorks.

We also found that scribe notes were on average 42% the length, in character count, of the notes extracted from the ASR transcript, which could also save time when used by organizers.

6 DISCUSSION

6.1 E-scribing is more inclusive for participants and time-efficient for organizers than existing approaches

We found that DeliberationWorks equipped scribes to capture and synthesize an average of 82% of the input shared by participants with only two hours of asynchronous training and preparation. The input captured was synthesized within 10 minutes after the deliberation was over. The input was labeled by the scribes with 77-94% accuracy (average across author, policy, and type labels) and 85% of the content captured was considered mostly accurate by two independent raters.

In the following sections, we discuss how the e-scribing approach compares to the three existing approaches for capturing and synthesizing participant input in face-to-face participatory policy-making deliberations.

6.2 Participant-generated input

The most popular digital technologies for supporting participatory policy-making aim to increase participation through online contributions. These digital technologies help increase participation because they make participation available to those who cannot attend face-to-face meetings. Yet, face-to-face deliberations are necessary for participatory policy-making because they allow participants to build trust, commitment to the process, and overall have a higher quality deliberation [1, 5, 6, 20].

To truly support inclusive participatory policy-making, technologies should be designed to support inclusive face-to-face deliberations. Compared to existing participant-generated input capture technologies [8, 22, 23, 40, 42, 43, 45], the main benefit of DeliberationWorks is that participants do not have to interact with technology at all to have their input captured and integrated into an online ecosystem, making participation more inclusive for participants who are unwilling or uncomfortable interacting with technology [44], those who have trouble expressing themselves in writing, or those that do not own digital devices with internet access. Based on our experience with the participatory budgeting process, most participants that attend face-to-face deliberations fall under at least one of these categories.

Nevertheless, e-scribing and participant-generated input naturally complement each other because they support different user groups: e-scribing supports inclusive face-to-face participation, while the participant-generated input approach supports participants who cannot attend face-to-face deliberations.

6.3 AI-based approaches

AI approaches promise to automate undesirable human tasks or process large amounts of information quickly, such as capturing and synthesizing participant input in participatory policy-making deliberations. Yet, AI is not ready to perform automated capture or synthesis because of the lack of accuracy and reliability [24, 28, 32].

Using existing algorithms, organizers still have to spend a significant amount of time fixing inaccuracies from transcripts and interpreting results of automated synthesis. We need further research to understand how much time exactly it would take organizers to capture and synthesize input using AI, but prior studies and our personal experience using AI-based systems, suggest it would take at least a few hours to days depending on how much data there is. State of the art commercial transcription software offers a series of recommendations to improve the accuracy of automated transcriptions, such as using external microphones, minimize background noise, avoid overlapping dialog [35]. However, these recommendations are not feasible in the context of face-to-face participatory policy-making without increasing the cost, such as by adding microphones to each participant or group. It is possible that even with extra hardware (e.g., microphones), ASR accuracy would not be high due to the noisy environment.

Compared to ASR technologies, DeliberationWorks scribes reduce the time that organizers need to synthesize the captured input from hours to minutes. This is because even if AI can generate a completely accurate transcription of the meeting, organizers still must go through the data to label participants' input. Another benefit of human scribes using DeliberationWorks suggested by our findings is that human scribes are able to capture intent, which an ASR technology would not be able to do.

E-scribing and AI-based approaches could potentially be combined to increase input capture accuracy and decrease synthesis time. AI-based technologies could help scribes increase their accuracy when using DeliberationWorks. For example, when scribes use DeliberationWorks and capture an input that looks like a question, an AI agent could show a pop-up that confirms with the scribe that their label is accurate. At the same time, e-scribing could also help AI increase its accuracy. For example, as scribes use DeliberationWorks to label participants' input, their data could be used to train machine learning models to automatically label scribes notes.

6.4 Scribe-based approaches

Scribe-based approaches are simple to implement and have already been used successfully at scale [e.g., 11, 19]. But without technology support, the cognitive load to capture and synthesize input might affect accuracy and completeness of their notes. Existing scribe-based approaches increase inclusion at face-to-face deliberations because they allow participants to share their input by simply having a discussion with their peers, but scribes currently are not equipped to synthesize discussions and save organizers time. This is because scribes are generally tasked to capture notes rather than labelling and categorizing them in a structured manner.

Digital technology can scribes do more than just capture notes to help organizers save time. As demonstrated by our work, by having technology that is simple to use, scribes are able to help organizers save time by doing some basic synthesis, such as by segmenting specific bits of input and labeling them almost in real time.

E-scribing also makes scribing more motivating for scribes. Anecdotally, when our research team deployed DeliberationWorks with student scribes, they were more motivated to use the tool as opposed to the off-the-shelf note-taking technology previously used. Offering scribes motivating tools is important because scribes are often volunteers and keeping volunteers engaged is a challenge to organizations running participatory policy-making processes [16].

6.5 Limitations and future work

This study has several limitations. First, we did not directly compare the e-scribing approach to existing approaches. To get a better understanding of how an AI-based approach compares to e-scribing, for example, future work should consider setting up experiments to directly compare existing approaches in terms of accuracy of input capture and time to train organizers to synthesize input.

We also focus on only one context of participatory budgeting. While this is one of the most critical contexts in which participants' input might influence budget allocations directly, there are many contexts in which capturing and synthesizing participants' input in real time might benefit both participants and organizers. Future work should incorporate e-scribing in contexts such as citizen assemblies with a diverse set of participants beyond students.

We learned that there was some significant variation between scribes' accuracy. Future work should consider improving the accuracy and completeness of scribes. For example, to make scribes more consistent without significantly increasing organizers time, the scribe training could include a short assessment of scribing ability or instruct scribes to actively clarify participants' input.

7 CONCLUSION

Citizens can increase openness, transparency, and accountability of institutions to address pressing social issues by taking part in participatory policy-making deliberations. Yet, for participants' contributions to influence policy outcomes, organizers must capture and synthesize participants' input. Existing approaches such as participant-generated input, artificial intelligence-based, and scribe-based, are not inclusive for participants or require too much time from organizers. In this paper, we present the design of e-scribing, a novel approach for capturing and synthesizing participant input from face-to-face participatory policy-making deliberations. To evaluate the e-scribing approach, we built DeliberationWorks, a digital deliberation technology that helps scribes capture input as participants and synthesize the discussion in real time. We deployed DeliberationWorks in two participatory budgeting deliberations. We found that, on average, 82% of the input was captured mostly accurately. Synthesis time was reduced to about 10 minutes and training time to one hour. Our findings suggest that e-scribing technologies can be used to support face-to-face deliberations to help make participation more inclusive for participants and save organizers time.

This work shows that e-scribing helps advance the goals of the digital governance by contributing the design and evaluation of a digital technology. We also contribute a new set of outcomes previously overlooked by the community that similar systems should consider when evaluated, such as training time. This work extends prior theories of inclusive participation in participatory policy-making without compromising inclusion or time required to scale the approach.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No. 2008450. We are grateful to Morgan Wu for her contributions to data analysis.

REFERENCES

[1] Chris Ansell and Alison Gash. 2008. Collaborative Governance in Theory and Practice. *Journal of Public Administration Research and Theory* 18, 4 (Oct. 2008), 543–571. https://doi.org/10.1093/jopart/mum032

[2] Miguel Arana-Catania, Felix-Anselm Van Lier, Rob Procter, Nataliya Tkachenko, Yulan He, Arkaitz Zubiaga, and Maria Liakata. 2021. Citizen Participation and Machine Learning for a Better Democracy. *Digital Government: Research and Practice* 2, 3 (July 2021), 27:1–27:22. https://doi.org/10.1145/3452118

[3] Ted Byrt, Janet Bishop, and John B. Carlin. 1993. Bias, prevalence and kappa. *Journal of Clinical Epidemiology* 46, 5 (May 1993), 423–429. https://doi.org/10.1016/0895-4356(93)90018-V

[4] Claudia Chwalisz. 2020. *Reimagining democratic institutions: Why and how to embed public deliberation.* Technical Report. OECD, Paris. https://doi.org/10.1787/056573fa-en

[5] Eric Corbett and Christopher A. Le Dantec. 2018. Going the Distance: Trust Work for Citizen Participation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173886

[6] Eric Corbett and Christopher A. Le Dantec. 2018. The Problem of Community Engagement: Disentangling the Practices of Municipal Government. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174148

[7] C. C. Crawford. 1925. The Correlation between College Lecture Notes and Quiz Papers. *The Journal of Educational Research* 12, 4 (Nov. 1925), 282–291. https://doi.org/10.1080/00220671.1925.10879600 Publisher: Routledge _eprint: https://doi.org/10.1080/00220671.1925.10879600.

[8] Jonathan Davies, Miguel Arana-Catania, and Rob Procter. 2022. Embedding digital participatory budgeting within local government: motivations, strategies and barriers faced. In *15th International Conference on Theory and Practice of Electronic Governance.* ACM, Guimarães Portugal, 98–104. https://doi.org/10.1145/3560107.3560124

[9] B. De Wever, T. Schellens, M. Valcke, and H. Van Keer. 2006. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education* 46, 1 (Jan. 2006), 6–28. https://doi.org/10.1016/j.compedu.2005.04.005

[10] Jeroen Delfos, Anneke Zuiderwijk, Sander Van Cranenburgh, and Caspar Chorus. 2022. Perceived challenges and opportunities of machine learning applications in governmental organisations: an interview-based exploration in the Netherlands. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance (ICEGOV '22).* Association for Computing Machinery, New York, NY, USA, 82–89. https://doi.org/10.1145/3560107.3560122

[11] Titiana-Petra Ertiö, Pekka Tuominen, and Mikko Rask. 2019. Turning Ideas into Proposals: A Case for Blended Participation During the Participatory Budgeting Trial in Helsinki. In *Electronic Participation (Lecture Notes in Computer Science)*, Panos Panagiotopoulos, Noella Edelmann, Olivier Glassey, Gianluca Misuraca, Peter Parycek, Thomas Lampoltshammer, and Barbara Re (Eds.). Springer International Publishing, Cham, 15–25. https://doi.org/10.1007/978-3-030-27397-2_2

[12] James S Fishkin. 2018. Deliberative polling. In *The Oxford handbook of deliberative democracy.* Oxford University Press, 315–328.

[13] Abraham E. Flanigan, Kenneth A. Kiewra, Junrong Lu, and Dzhovid Dzhuraev. 2023. Computer versus longhand note taking: Influence of revision. *Instructional Science* 51, 2 (April 2023), 251–284. https://doi.org/10.1007/s11251-022-09605-5

[14] Abraham E. Flanigan and Scott Titsworth. 2020. The impact of digital distraction on lecture note taking and student learning. *Instructional Science* 48, 5 (Oct. 2020), 495–524. https://doi.org/10.1007/s11251-020-09517-2

[15] Hahrie Han. 2014. *How Organizations Develop Activists: Civic Associations and Leadership in the 21st Century.* Oxford University Press. Google-Books-ID: rOESDAAAQBAJ.

[16] Soo-Hye Han, William Schenck-Hamlin, and Donna Schenck-Hamlin. 2015. Inclusion, Equality, and Discourse Quality in Citizen Deliberations on Broadband. *Journal of Deliberative Democracy* 11, 1 (May 2015). https://doi.org/10.16997/jdd.220 Number: 1 Publisher: University of Westminster Press.

[17] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing Community-Based Collaborative Design: Towards More Equitable Participatory Design Engagements. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 216:1–216:25. https://doi.org/10.1145/3359318

[18] Christina N. Harrington, Katya Borgos-Rodriguez, and Anne Marie Piper. 2019. Engaging Low-Income African American Older Adults in Health Discussions through Community-based Design Workshops. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300823

[19] Janette Hartz-Karp. 2005. A Case Study in Deliberative Democracy: Dialogue with the City. *Journal of Deliberative Democracy* 1, 1 (April 2005). https://doi.org/10.16997/jdd.27 Number: 1 Publisher: University of Westminster Press.

[20] Janette Hartz-Karp and Brian Sullivan. 2014. The Unfulfilled Promise of Online Deliberation. *Journal of Deliberative Democracy* 10, 1 (June 2014). https://doi.org/10.16997/jdd.191 Number: 1 Publisher: University of Westminster Press.

[21] Mahmood Jasim, Enamul Hoque, Ali Sarvghad, and Narges Mahyar. 2021. CommunityPulse: Facilitating Community Input Analysis by Surfacing Hidden Insights, Reflections, and Priorities. In *Designing Interactive Systems Conference 2021 (DIS '21).* Association for Computing Machinery, New York, NY, USA, 846–863. https://doi.org/10.1145/3461778.3462132

[22] Mahmood Jasim, Pooya Khaloo, Somin Wadhwa, Amy X. Zhang, Ali Sarvghad, and Narges Mahyar. 2021. CommunityClick: Capturing and Reporting Community Feedback from Town Halls to Improve Inclusivity. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 213:1–213:32. https://doi.org/10.1145/3432912

[23] Ian G. Johnson, Alistair MacDonald, Jo Briggs, Jennifer Manuel, Karen Salt, Emma Flynn, and John Vines. 2017. Community Conversational: Supporting and Capturing Deliberative Talk in Local Consultation Processes. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17).* Association for Computing Machinery, New York, NY, USA, 2320–2333. https://doi.org/10.1145/3025453.3025559

[24] Saba Kawas, George Karalis, Tzu Wen, and Richard E. Ladner. 2016. Improving Real-Time Captioning Experiences for Deaf and Hard of Hearing Students. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '16).* Association for Computing Machinery, New York, NY, USA, 15–23. https://doi.org/10.1145/2982142.2982164

[25] Kenneth A. Kiewra. 2016. Note Taking on Trial: A Legal Application of Note-Taking Research. *Educational Psychology Review* 28, 2 (June 2016), 377–384. https://doi.org/10.1007/s10648-015-9353-z

[26] Hélène Landemore. 2020. *Open Democracy.* https://press.princeton.edu/books/hardcover/9780691181998/open-democracy

[27] J. R. Landis and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (March 1977), 159–174.

[28] Walter Lasecki, Christopher Miller, Adam Sadilek, Andrew Abumoussa, Donato Borrello, Raja Kushalnagar, and Jeffrey Bigham. 2012. Real-time captioning by groups of non-experts. In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12).* Association for Computing Machinery, New York, NY, USA, 23–34. https://doi.org/10.1145/2380116.2380122

[29] Seunghoo Lim and Youngmin Oh. 2016. Online Versus Offline Participation: Has the Democratic Potential of the Internet Been Realized? Analysis of a Participatory Budgeting System in Korea. *Public Performance & Management Review* 39, 3 (July 2016), 676–700. https://doi.org/10.1080/15309576.2016.1146553 Publisher: Routledge _eprint: https://doi.org/10.1080/15309576.2016.1146553.

[30] Narges Mahyar, Diana V. Nguyen, Maggie Chan, Jiayi Zheng, and Steven P. Dow. 2019. The Civic Data Deluge: Understanding the Challenges of Analyzing Large-Scale Community Input. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19).* Association for Computing Machinery, New York, NY, USA, 1171–1181. https://doi.org/10.1145/3322276.3322354

[31] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 72:1–72:23. https://doi.org/10.1145/3359174

[32] Moira McGregor and John C. Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17).* Association for Computing Machinery, New York, NY, USA, 2208–2220. https://doi.org/10.1145/2998181.2998335

[33] Matthew B. Miles, A. Michael Huberman, and Johnny Saldaña. 2014. *Qualitative Data Analysis.* SAGE Publications. Google-Books-ID: 3CNrUbTu6CsC.

[34] Simon Niemeyer and John S. Dryzek. 2007. The Ends of Deliberation: Meta-consensus and Inter-subjective Rationality as Ideal Outcomes. *Swiss Political Science Review* 13, 4 (2007), 497–526. https://doi.org/10.1002/j.1662-6370.2007.tb00087.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1662-6370.2007.tb00087.x.

[35] Otter.ai. 2022. Transcription accuracy FAQ. https://help.otter.ai/hc/en-us/articles/360048322533-Transcription-accuracy-FAQ

[36] Fernando Pinto, Marie Anne Macadar, and Gabriela Viale Pereira. 2022. The potential of eParticipation in enlarging individual capabilities: a conceptual framework. *Information Technology for Development* 0, 0 (2022), 1–23. https://doi.org/10.1080/02681102.2022.2136129 arXiv:https://doi.org/10.1080/02681102.2022.2136129

[37] Daniela K. Rosner, Saba Kawas, Wenqi Li, Nicole Tilly, and Yi-Chen Sung. 2016. Out of Time, Out of Place: Reflections on Design Workshops as a Research Method. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1131–1141. https://doi.org/10.1145/2818048.2820021

[38] Anwar Shah. 2007. *Participatory Budgeting*. World Bank Publications. Google-Books-ID: Y1WQYgC9JNEC.

[39] Stan W. Smith. 2010. An experiment in bibliographic mark-up: Parsing metadata for XML export. In *Proceedings of the 3rd. annual workshop on Librarians and Computers (LAC '10, Vol. 3)*, Reginald N. Smythe and Alexander Noble (Eds.). Paparazzi Press, Milan Italy, 422–431. https://doi.org/99.9999/woot07-S422

[40] Nick Taylor, Justin Marshall, Alicia Blum-Ross, John Mills, Jon Rogers, Paul Egglestone, David M. Frohlich, Peter Wright, and Patrick Olivier. 2012. Viewpoint: empowering communities with situated voting devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1361–1370. https://doi.org/10.1145/2207676.2208594

[41] Nivek K. Thompson. 2012. Participatory budgeting - the Australian way. *Journal of Deliberative Democracy* 8, 2 (Dec. 2012). https://doi.org/10.16997/jdd.145 Number: 2 Publisher: University of Westminster Press.

[42] Sunny Tian, Amy X. Zhang, and David Karger. 2021. A System for Interleaving Discussion and Summarization in Online Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 241:1–241:27. https://doi.org/10.1145/3432940

[43] Vasilis Vlachokyriakos, Rob Comber, Karim Ladha, Nick Taylor, Paul Dunphy, Patrick McCorry, and Patrick Olivier. 2014. PosterVote: expanding the action repertoire for local political activism. In *Proceedings of the 2014 conference on Designing interactive systems (DIS '14)*. Association for Computing Machinery, New York, NY, USA, 795–804. https://doi.org/10.1145/2598510.2598523

[44] Jenny Waycott, Frank Vetere, Sonja Pedell, Amee Morgans, Elizabeth Ozanne, and Lars Kulik. 2016. Not For Me: Older Adults Choosing Not to Participate in a Social Isolation Intervention. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 745–757. https://doi.org/10.1145/2858036.2858458

[45] Amy X. Zhang and Justin Cranshaw. 2018. Making Sense of Group Chat through Collaborative Tagging and Summarization. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 196:1–196:27. https://doi.org/10.1145/3274465