

DevCAFE 1.0: A participatory platform for assessing development initiatives in the field

Brandie Nonnecke^{a,w}, Sanjay Krishnan^{b*}, Jay Patel^b, Mo Zhou^c, Laura Byaruhanga^d, Dorothy Masinde^c, Maria Elena Meneses^f,
Alejandro Martin del Campo^f, Camille Crittenden^a, Ken Goldberg^{a,b,c}

^aCITRIS Connected Communities Initiative, UC Berkeley, ^bElectrical Engineering and Computer Science, UC Berkeley,

^cIndustrial Engineering and Operations Research, UC Berkeley, ^dUganda Program, Iowa State University

^eGlobal Resource Systems, Iowa State University, ^fTecnológico de Monterrey

*These authors contributed equally to this work, ^wCorresponding Author

Abstract—The design and assessment of development initiatives is increasingly participatory, where decision makers consider feedback from affected populations. While digital data collection facilitates faster and more reliable analysis, existing data collection tools are not optimized for unstructured qualitative (textual) data and peer-to-peer participant collaboration. In this paper, we propose a system called the Development Collaborative Assessment and Feedback Engine version 1.0 (DevCAFE), a customizable participatory assessment platform that collects and integrates quantitative assessment, qualitative feedback and peer-to-peer collaborative filtering. DevCAFE incorporates a library of statistical analyses for researchers to quickly identify quantitative and qualitative trends while collecting field data. DevCAFE can run on any mobile device with a web-browser and can work with or without Internet connectivity. We present results from two pilot projects: (1) 137 participants evaluating family planning education trainings at three Nutrition Education Centers in rural Uganda, and (2) 4,518 participants evaluating policy priorities for elected leaders in the June 2015 Mexico mid-term elections. DevCAFE collected over 19,000 peer-to-peer ratings of 336 submitted ideas. Feedback gathered through DevCAFE enabled targeted reforms to the family planning efforts in Uganda and the need for increased government attention to public safety in Mexico. Case studies and interactive demos are available at: <http://opinion.berkeley.edu/devcafe/>

Keywords—collaborative filtering, development assessment, participatory assessment, principal component analysis

I. INTRODUCTION

A number of recent studies suggest that including affected communities in the design and implementation of development initiatives improves their efficacy [47-50]. Participatory methods for assessment are an established practice in development policy and both the World Bank and United Nations have specified standards for their application [45-47]. However, the primary bottleneck is collecting sufficient data from the field, especially from remote regions, and analyzing this data in a timely manner. Digital survey tools such as Open Data Kit (ODK) and SMS-based mobile frameworks [2,3] are revolutionizing field data collection. These tools use a digital survey design to reduce the time-consuming and error-prone process of converting handwritten notes, paper forms, and audio recordings into structured digital data [8].

Digital survey tools are useful for experimental designs and confirmatory analysis as they mitigate sources of bias [5]. However, development participants, practitioners and researchers are often interested in participatory, exploratory analysis to discover patterns, problems and issues of interest while in the field [16]. The

This work was supported by the Blum Center for Developing Economies and the Development Impact Lab (USAID Cooperative Agreement AID-OAA-A-12-00011), part of the USAID Higher Education Solutions Network; UC Berkeley’s Algorithms, Machines, and People Lab; and the UC CITRIS Connected Communities Initiative.

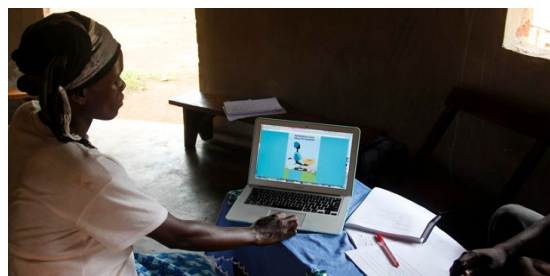


Fig. 1. The DevCAFE v1.0 platform was applied to assess family planning education programs in rural Uganda and to evaluate government performance in Mexico. Over 4,500 participants contributed responses through the platform.

exploratory analysis phase is inherently less structured, requiring more qualitative questions (i.e., open-ended text). In this setting, scalable and timely analysis of qualitative data (e.g., with NLP) can be challenging due to local language and domain-specificity. During the exploratory phase, researchers may need to collect more quantitative data (e.g., demographics) to test relevant hypotheses.

Data collection for exploratory analysis has two major data filtering challenges: (1) qualitative data filtering where researchers must quickly identify relevant open-ended responses, and (2) quantitative data filtering where researchers must identify which quantitative variables (or groups of variables) correlate significantly with demographics or qualitative features. In our initial version of the Development Collaborative Assessment and Feedback Engine (DevCAFE), we address these two challenges. DevCAFE is a mobile-optimized assessment platform that enables rapid exploratory analysis. We address qualitative data filtering challenge through collaborative filtering, where participants evaluate each others’ responses as well as providing their own. The system quickly extracts ideas that resonate with the group through sampling and ranking [4,6]. Quantitative data filtering challenges are addressed through dimensionality reduction, where we find features (i.e., linear combinations of responses) that best differentiate participants through Principal Component Analysis (PCA). These features can be correlated with demographics and qualitative responses and tested for significance.

DevCAFE features a visual, interactive interface to mitigate the tedium of other peer-to-peer evaluation systems where participants sift through long lists. Instead, DevCAFE presents a 2D map showing a small sample of ideas to evaluate. This visual interface is designed in HTML5, works across platforms and is extensible to different languages. Evaluations and quantitative data are collected through visual modalities such as sliders and keypads that can be adapted to different local settings.

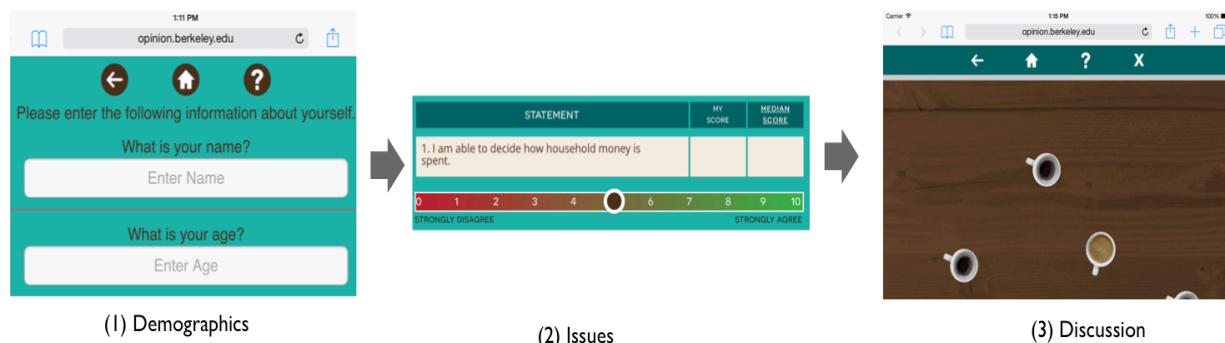


Fig. 2. Screenshots of the DevCAFE platform. Participants first respond to a series of demographic questions. Then, they provide quantitative responses on key issues with a scale from 0 to 10 (“Strongly Disagree” to “Strongly Agree”). Finally, they enter an open-ended discussion (in this version visualized as a metaphorical coffee table), where they can click on a mug and read and evaluate other participants’ responses.

In addition, many development studies are in locations with intermittent or non-existent Internet connectivity. DevCAFE supports the mobility features of existing digital assessment platforms: (1) DevCAFE can be loaded on any device with an HTML5-compatible browser, (2) it does not require Internet connectivity once loaded, and (3) if connectivity is available, the system will synchronize with its server. DevCAFE is a flexible platform that can be modified to fit the needs of a participatory assessment application. Researchers can select demographic, quantitative and qualitative questions. We implement many of the salient features of tools like ODK while adding additional support for peer-to-peer interactions, visualization, and statistical analysis.

To evaluate this initial version we describe two case studies:

Uganda Case Study. We applied DevCAFE for use in Kamuli, Uganda in June 2014 to assess effectiveness of family planning trainings at three Nutrition Education Centers (Centers). Responses were collected from 137 women who attended these trainings.

Mexico Case Study. We applied a Spanish-language DevCAFE to evaluate the Mexican Government’s performance on timely policy issues. We collected data from April 2015 to June 2015 from 4,518 participants from throughout the country.

II. RELATED WORK

Digital data collection tools for humanitarian programs have been well studied [13]. The most comprehensive toolkit is the Open Data Kit (ODK) project [7,8,21]. ODK is an Android-based data collection tool, which uploads collected data to a central server. It is resilient to network connectivity problems and can export data in many standard research formats. It further integrates with sensor, image, and voice data and has meta-tools, allowing researchers to easily design surveys and survey protocols. While the functionalities of ODK and DevCAFE overlap, we optimize our software for a different application. We target collaborative participant-centric problem exploration and issue discovery, and manage this data longitudinally. Participants not only contribute data, but also receive instant feedback and can evaluate responses provided by others. To implement interactive applications in an intermittent network, we apply results from the database literature such as transactions and logging. These two features allow us to operate in unreliable networks with less fear of losing data.

One alternative to ODK and related tools has been SMS-based data collection such as FrontlineSMS and RapidSMS [1,21,28]. FrontlineSMS is an open-source platform for surveying communities with limited Internet access but with SMS coverage. As with ODK, the system allows for collecting responses to simple forms via SMS,

but does not enable interaction between participants. While FrontlineSMS is primarily used to gather data for short surveys, RapidSMS can be used to gather more structured data on a larger scale. RapidSMS is an open-source platform built with Python and Django. The use of SMS for data collection has been proven successful in a variety of development research contexts; however, research has shown that data collected through SMS may not be as accurate as data collected through the use of interactive voice recognition systems [30].

Within developing country contexts, SMS data collection capabilities may be diminished due to literacy and language barriers among target populations. One solution to this problem has been the use of interactive voice recognition systems (IVRs). ODK Voice is one tool that enables surveys to be recorded in the local language or dialect [21]. To complete the survey, participants respond to an automated phone call or call a central phone number. Participants then provide answers to survey questions by entering numbers in response to audio-recorded questions or by recording qualitative answers (ibid.). While extremely useful, such approaches are still limited in the types of data they can collect. A related point of ensuring the quality of collected data has also been studied [5, 9, 29].

Participatory action research (PAR) engages participants as active contributors in the research process [39]. Participants are encouraged to identify problems faced within the community, interpret research results, and identify potential applications of these results to address local problems. The bottom-up approach of PAR is particularly useful in development-based research where interventions must draw on local knowledge to better understand factors affecting successful implementation. This theoretical model argues that when subjects actively participate rather than passively provide data it can yield more informative results in democratic processes, strategic planning and ideation.

DevCAFE builds on our prior work in designing citizen engagement platforms. In the OpinionSpace project [17], our results suggested that 2D visual interfaces for presenting ideas increased engagement. We have further explored best practices for online civic engagement platforms [52], bias mitigation due to social influences [53], and applying similar ideas to large college course evaluations [54]. The development setting poses a number of new challenges including: (1) Internet connectivity, (2) language and cultural differences, and (3) rapid field analysis.

III. DEVCAFE PLATFORM

A. Overview

DevCAFE is a new assessment platform that can directly engage citizens and communities from developing regions to collectively assess conditions, needs and outcomes related to development projects. One key focus of the platform is to collect longitudinal data, and track changes over time. In contrast to polls and surveys traditionally used by development researchers, DevCAFE is user-facing and provides direct, immediate visual feedback to participants, allowing them to instantly understand where they stand on key issues in relation to other participants and encourage open-ended / qualitative responses that can enhance interaction between participants. We envision DevCAFE as a grassroots, user-friendly platform that allows the wisdom of crowds to identify, highlight and display patterns, trends and insights as they emerge.

DevCAFE features three core stages (Figure 2). In the first stage, participants provide demographics. In the second stage, participants provide quantitative assessments of key issues of importance in the study. Finally, participants provide a qualitative (textual) response to an open-ended question and evaluate the importance of others' responses. All responses are anonymous. The design can easily be adapted to accommodate different cultural settings and technical requirements.

B. Demographic Questions

DevCAFE is designed to collect direct input on demographics from participants or researchers during an interview. The first stage asks participants to answer a series of demographic questions (e.g., age, gender, religion). Questions about the relationship between the participant and the development program can be asked (e.g., how many years have you been involved in this development program?). The exact questions and the domain of their responses are configurable.

C. Issue Assessment Questions

The second stage encourages participants to answer quantitative assessment questions. We frame these questions to be answered with a slider on a numerical scale (e.g., 0-10, continuous 0 to 1, etc.). We anchor the ends of the slider with descriptors such as "Strongly Disagree" to "Strongly Agree." The text, scale and properties of the sliders are configurable. Example questions include: "Rate from strongly disagree to strongly agree: I am able to decide how household money is spent." Furthermore, participants have the option

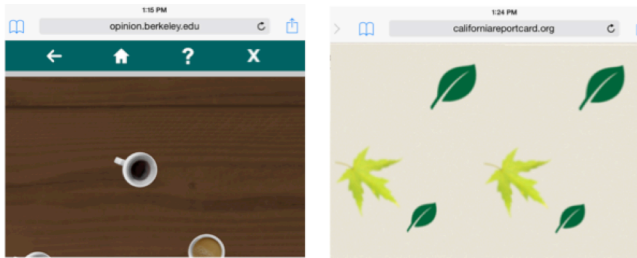


Fig. 4. The DevCAFE v1.0 visualization can be adapted to many different cultural settings while preserving the same workflow. In this figure, we visualize participants' ideas with different objects.

to skip any questions they do not feel comfortable answering. Upon answering an assessment question, the participant is shown the median response over all prior participants. We use the feedback as an incentive to encourage participants to provide more information.

The assessment questions serve two purposes. First, they provide researchers with quantitative data to better understand the diversity of the population being assessed and the effectiveness of interventions. Second, the questions reveal to the participants where they stand with respect to others. Participants instantly know if they largely agree with the median or if their responses are more extreme than others in the system.

D. Open-Ended Questions

In this stage, there is a primary open-ended discussion question. For example, "How can the Nutrition Education Centers be improved and why?" Participants evaluate the importance of others' suggestions by assigning a score from 0 to 10 using a slider (e.g., "Not at All Important" = 0 to "Extremely Important" = 10) (Figure 3). After reading and rating two suggestions, participants are invited to enter their own suggestion. We designed the system this way to allow participants to learn from others in the system and think about the question before responding.

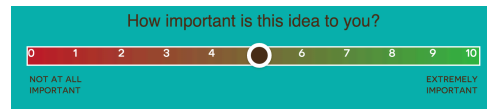


Fig. 3. Participants rate each other's ideas on importance. Peer-to-peer ratings are an important part of DevCAFE v1.0's solution to the potential bottleneck of analyzing qualitative data.

E. Discussion Visualization

List-based interfaces often lead to tedium, can be difficult to use on mobile devices, and can lead to a "winner takes all" dilemma where popular suggestions are prioritized over new or novel suggestions. To mitigate these negative effects, each participant's suggestion is visualized as an object (e.g., cup or sphere) on a 2D map (Figure 4). The interface can be adapted to different cultural norms. In [17], the authors showed that this interface led to improved discussion quality. When participants select an object, they can read and evaluate another participants' suggestion. In future work, we are exploring supporting audio recordings of qualitative data because participants may have varying levels of numerical and textual literacy.

Objects in closer proximity represent participants who responded to the quantitative issue assessment questions similarly. This allows participants to immediately see how people similar to them feel about the discussion question. We apply Principal Component Analysis (PCA) to place objects on the map with the participant at the center. In the first step, we associate each participant with a k -dimensional vector; each entry corresponds to one response to the assessment questions. We then apply PCA to the set of vectors and the algorithm returns a two dimensional coordinate (x,y) associated to each participant. This point corresponds to the top 2 eigenvectors of the covariance matrix, revealing the top two principal components that explain the most variance. We then center the visualization on the participant's (x_p, y_p) position, and then arrange the objects in the new coordinate space.

It would be infeasible to show participants all of the suggestions simultaneously. We strategically select which suggestions to display to participants. We calculate the standard error when evaluating each suggestion; this measures how uncertain we are about the mean evaluation. Using the standard error, we weight each response, and use this weight to prioritize which suggestions are shown. We select the suggestions in a randomized way where suggestions with a higher weight (standard error) are more likely to be selected.

F. Database and Network Architecture

The architecture of DevCAFE is a client-server model. An HTML5 client communicates with a DevCAFE server. This server is a Django-based application, which can run both on internetworks and local area networks. For example, we can host the server on a public webserver and then the application can be visited from anywhere with network connectivity and any device with a browser. In contrast, we can also host the server on a laptop and access the application from any device on the laptop's local network (including the laptop itself) without Internet access. DevCAFE can be deployed in mixed networks, combining local networks and communicating via the Internet. This is useful since field research may be conducted in a developing region with limited Internet access, but collaborators may be in developed countries with available connectivity. The same application deployed in the field can be hosted on the Internet and asynchronously updated with new field data.

In Figure 5, we illustrate the different network architectures supported by DevCAFE. One drawback to the interactive design of the application is what is sometimes referred to as "chattiness" of the communication protocol. Since we communicate with the server in many small transactions this can be inefficient when network latencies are large. We took steps to combine small transactions into larger ones; however, we defer full treatment of this subject to future work.

Intermittent connectivity between the server and the client also poses a challenge for the database programming. We need to ensure that all data is eventually written to the server and that a loss in connectivity does not result in lost data. We address this problem with a technique called transaction logging [32]. We log the attempt and its parameters before every communication attempt. Then, when communications start to fail we mark the point in the log of the last successful attempt. Finally, when network connectivity is available the log is replayed and these operations are committed to the database.

To achieve this we have to be careful about what types of database operations we allow. We designed our program to have insert-only interactions with the database; we do not modify or delete existing records. When a record is updated, it is marked as old and a new record is created with the updated value. This allows us to reconcile transactions that arrive after-the-fact by checking their timestamps. There is no danger that the data in question has been deleted.

IV. RESEARCHER INTERFACE

Next, we discuss how researchers can use the DevCAFE interface to configure the platform, and discover emerging trends in the data.

A. Configuration and Customization

We provide a web-based interface to change all text and visuals in the platform. The interface, which we call the *Admin Panel*, allows the researcher to change text, visuals and settings of the system at any time.

B. Translation

Often the research team collecting data relies on local collaborators for translation. As a result, analysis or validation of the data is difficult. When available, DevCAFE uses Google Translate to automatically translate submitted responses. This translation gives the researcher a cursory idea of what the participant is saying. The interface also supports human-in-the-loop editing and refinement of the translations. In our current implementation, we support a bilingual

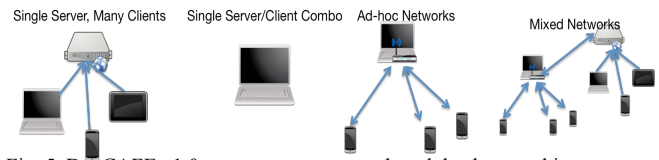


Fig. 5. DevCAFE v1.0 can use many network and database architectures.

system and hope to expand our interfaces to support many-to-many translations.

C. Tools for Statistical Analysis

1) *Quantitative Data and Demographics*: As in our visualization, our main methodology is to apply dimensionality reduction to the issue assessment questions and identify a small number of "factors" that differentiate participants. We then correlate these factors to the demographics, to suggest which demographic variables relate to which factors. We can then test statistical significance of these relationships. Formally, PCA finds a set of uncorrelated features (linear combinations of observed variables) with maximum variance. In other words, it chooses axes that differentiate (spread out) data points the best. The interpretation of PCA features is they are latent (unobserved) factors. For example, in psychology, applying PCA personality surveys results in the 5 main personality types [15]. Likewise, in political science, applying PCA to political position surveys results in one axis representing social views and one representing fiscal views [22]. In our context, PCA provides a computationally efficient way to explore the quantitative data, and understand which questions are correlated and which questions are most informative. Using these PCA features, we can then correlate these features to the textual data or demographic variables, allowing us to measure the correlation coefficients between a demographic variable and the principal components. This allows for a quick test to see if that demographic variable is correlated with assessment responses.

2) *Qualitative Data*: To analyze the qualitative data, we combine the peer-to-peer evaluations into a reputation model, which scores each response. This score also incorporates the variance in the evaluations and the sample size. To calculate the score, we first take the mean evaluation \mathbf{g} . We then calculate the approximate 95% confidence interval of the \mathbf{g} using the standard error of $\mathbf{g} \pm 1.96 * SE(\mathbf{g})$. We rank the comments by the lower bound $\mathbf{g} - 1.96 * SE(\mathbf{g})$. This approach, also called Wilson Score Ranking, has been applied to other problems in crowdsourcing [6].

V. CASE STUDY AND RESULTS

In the following, we describe two applications of DevCAFE and how insights from these applications inform future work.

A. Uganda Case Study

From June 16th -19th, 2014, a member of our team conducted a series of interviews with 137 women at Nutrition Education Centers (Centers) in the Kamuli District, Uganda using the DevCAFE software. The Centers provide daily nutrient-dense meals to mothers and children and trainings in health and sanitation, family planning, agricultural productivity, and income-generating activities [42]. The effectiveness of improved agricultural practices and skills trainings have been studied within this population [27]. Our work extends evaluation to effectiveness of family planning trainings.

The Centers highlight maternal health as a key priority and offer family planning training for women [42]. In this project, we used DevCAFE to collect feedback on the quality of the family planning trainings offered and to understand the relationships between

demographic variables and participants' opinions toward family planning methods. The interface and data are available at (<http://opinion.berkeley.edu/uganda/mobile/>).

A researcher on our team installed the software on an Apple MacBook Air and collected data through DevCAFE on the laptop with the assistance of a translator conversant in English and the local language, Lusoga. DevCAFE included demographic questions (e.g., age, education level, etc.), six quantitative assessment questions, and an open-ended qualitative question. DevCAFE asked participants to answer the following quantitative questions on a scale from 0-10, representing strongly disagree to strongly agree:

- I am able to decide how household money is spent. (*Decision Making Autonomy*)
- My husband supports the use of family planning. (*Husband Support*)
- I openly discuss family planning with my husband. (*Open Discussion*)
- I am afraid that using an oral contraceptive pill will cause me to never have children in the future even after I stop taking the oral contraceptive pill. (*Fear Contraceptive Pill*)
- I am afraid that a coil (*Intrauterine Device*) will cause me to never have children in the future even after the coil is removed. (*Fear Intrauterine Device*)
- I am afraid that receiving the birth control injection will cause me to never have children in the future. (*Fear Injection*)

In addition to these questions, we asked a qualitative question about how participants would improve the Centers.

1) *Quantitative Data:* We first describe how PCA can be applied to identify important correlations. We apply PCA to the six quantitative assessment questions. In Table 1, we show the first two principal components. The first principal component represents questions relating to the participants' level of autonomy: monetary autonomy, support of family planning methods and degree she can openly discuss family planning with her husband. The questions most aligned with the second component, fear of 3 types of birth control: injection, intrauterine coil and oral contraceptive. There are two scores for each participant: one for the autonomy factor and one for the fear factor. This is a linear combination of the quantitative responses based on the weights below. These scores can tell us a lot about the participants. For example, a high household factor score indicates autonomy and support for family planning at home, but a high fears score indicates reservations about different contraceptive methods. We can use these scores to test correlations between demographic variables and user responses. By finding these scores automatically, we significantly help researchers find trends in new data sources where they may not have intuition on differentiating factors.

TABLE 1 PCA Factors for Uganda Case Study. Factor 1 is most aligned with Autonomy and Factor 2 is most aligned with Fears.

Issue	Factor 1 (Autonomy)	Factor 2 (Fears)
Decision Autonomy	0.354	0.001
Husband Support	0.287	-0.092
Open Discussion	0.451	-0.006
Fear Contraceptive Pill	-0.012	0.564
Fear Intrauterine Device	0.078	0.641
Fear Injection	-0.102	0.402

The factors can be correlated with the demographic data. Statistical analysis enabled through DevCAFE revealed the following insights:

(1) The autonomy feature was strongly correlated with age, implying that older women tend to make more of the household monetary decisions and were more open in discussing family planning with their husbands ($r=0.198, p<0.001$). (2) The center that the participant visited was very predictive of fears of contraceptive methods. For example, those who visited the Naluwoli Center were less likely to indicate fears of any specific contraceptive method ($r=0.254, p<0.001$). The Naluwoli Center is the longest-running Nutrition Education Center in the district. Due to its longer presence in the community, the Center has likely established credibility within the community.

2) *Qualitative Data:* In the first case study, we were unable apply peer-to-peer evaluation of ideas due to literacy and language barriers. We evaluate this aspect in further detail in the next case study, and include a discussion of the lessons learned, which we are addressing in future work. Nonetheless, to demonstrate the value of qualitative data, the following textual ideas were chosen based on their prevalence and novelty compared to all other suggestions:

“It is important to mobilize husbands to also attend family planning trainings.”

“We need a water well at the Center. We travel too far to get water for the Center.”

“We walk too far to get to the Center. More Centers should be opened.”

3) *Results and Lessons Learned:* As a result of the feedback enabled through DevCAFE, the Centers have implemented changes to increase the effectiveness of its programs in the Kamuli District. First, the Centers have increased from 3 to 8 to decrease the amount of time participants must travel to attend trainings. Second, the Centers have invested in the construction of water wells at the Centers to reduce time spent collecting water. Third, one of the Centers will serve as a test location for the inclusion of men in family planning trainings, responding to the women's requests that men be included in family planning trainings.

This case study suggests a number of important considerations when developing a system like DevCAFE. Making participatory assessment agnostic to language and accommodating of differing literacy levels is a significant challenge. Ideally, we would like the women themselves to provide information directly on the platform without a translator or researcher to collect the data. In future work, we are exploring visual- and audio-based interfaces with tablets to facilitate such interactions.

B. Mexico Case Study

In preparation for the June 2015 mid-term elections in Mexico, México Participa (<http://mxparticipa.org>) was launched in mid-April 2015 in collaboration with faculty at Tec de Monterrey and the Mexico National Electoral Institute. The DevCAFE instance was fully translated into Spanish. DevCAFE included demographic questions (e.g., age, education level, etc.), six quantitative assessment questions, and an open-ended qualitative question. DevCAFE asked participants to answer the following quantitative questions on a scale from 0-10, representing strongly disagree to strongly agree:

- There are opportunities to have a job with Social Security (*Social Security*)
- Children have access to quality public education. (*Public Education*)
- I have access to good public health services. (*Public Health*)

- My community is safe and secure. (*Security*)
- My community is free of violence. (*Violence*)
- My local government is free of corruption. (*Corruption*)

Participants also were asked an open-ended question where they could suggest an idea that deserves increased national priority. In this case study, participants also evaluated each other’s suggestions. México Participa was announced in over 100 Mexican newspapers and radio stations in the 6 weeks prior to the mid-term elections. Over 200 Twitter and Facebook posts were sent by our team to the 64 candidates running in the mid-term election to encourage participation among their constituents.

In comparison to the previous case study, this case study was deployed over the entire country and collected data from a diverse set of participants. México Participa had 4,518 participants who assigned 21,632 grades, provided 19,077 peer-to-peer ratings, and suggested 336 ideas. Participants accessed México Participa through a variety of devices, including Desktop (70.5%), Smartphone (18.8%), Tablet (10.7%). Due to our responsive HTML5 design, our system was able to work across these devices.

An important question about México Participa is its representativeness with respect to the population. Even with the traditional media and social media campaign, acquiring rural participants was a challenge. While all of the states were represented, the Mexico City federal district and the state around the capital (State of Mexico) accounted for 45% of participants while only accounting for 21% of the country’s population. Furthermore, most of the participants in México Participa were between 18 to 34 years old. We compare these age demographics to the demographics of Mexico collected in the 2010 Mexican Census [51] in Figure 6. On the other hand, the gender breakdown (51% Male vs. 48% Female) was even with statistically insignificant differences.

1) *Quantitative Data:* We apply PCA to the six quantitative questions and correlate the factors to demographics data. In Table 2, we show the top two PCA factors. For the first factor all of the weights are positive. We phrased the questions such that a high score rating implies a positive assessment of the Mexican government. Thus, this factor corresponds to general “satisfaction” with the government’s performance. The second factor assigns a negative weight to Social Security, Public Education, and Health while positive weights to the questions about Security, Violence and Corruption. This factor suggests there may be participants satisfied with the government’s progress on social services but dissatisfied with government progress on public safety and corruption.

TABLE 2. PCA Factors for México Participa. Factor 1 corresponds to general satisfaction with the government's progress and factor 2 shows a separation between social service performance and public safety.

Issue	Factor 1 (Satisfaction)	Factor 2 (Social Services vs. Safety)
Social Security	0.307	-0.525
Public Education	0.368	-0.436
Public Health	0.390	-0.341
Security	0.472	0.156
Violence	0.453	0.333
Corruption	0.434	0.530

The factors can be correlated with the demographic data, and list the statistically significant insights. We found a significant negative correlation between participant Age and Satisfaction ($r = -0.112$, $p < 0.001$). Older participants provided more negative assessments of the Mexican government in general than younger ones. This insight highlights one of the benefits of the multivariate analysis in DevCAFE. If we simply reported the mean agreement on each quantitative question, we would overestimate the means since older participants are less represented (i.e., age breakdown in Figure 6). By understanding how different combinations of questions correlate with demographics, we get a richer understanding of the relationships between all variables.

2) *Qualitative Data:* We collected 19,077 peer-to-peer ratings on 336 submitted ideas about problems facing Mexico. In this case study, we

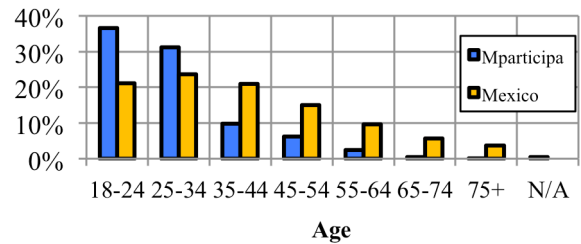


Fig. 6. The age breakdown of Mexico Participa vs. the Mexican Population. A majority of the participants are young and are overrepresented w.r.t the population.

had an opportunity to analyze the peer-to-peer rating system. Even if participants did not submit ideas, they still rated others’ ideas. 53% of participants provided at least one peer-to-peer rating. 14% rated at least 10 others, and 3% rated at least 20 others. We did not discover any significant correlations between demographics (age and location) and the number of peer-to-peer ratings or idea submissions.

A number of participants reiterated the initial six assessment questions in their ideas, e.g.,

“There are three problems that the government of Mexico needs to urgently fully resolve: corruption, insecurity and education”

We applied our ranking model to extract the highest-rated participant responses that were distinct from the original six. We present the three of highest-rated ideas:

“Insecurity and corruption, unemployment or low-wage employment, all lead to a lack of opportunities for youth and adults [...]”

“The bureaucracy in the National Anti-Corruption System will only grow, we must address the root causes of corruption among public officials.”

“The lack of credibility of our leaders. We need to trust the people who wish to govern. Not only do we need to eliminate corruption, we need fully trained leaders for public administration and true vocation of service.”

3) *Results and Lessons Learned:* Feedback gathered through México Participa was shared with newly elected leaders and incumbents through the Mexico National Electoral Institute. Across Mexico, participants expressed the need for increased government efforts to mitigate causes of corruption, strengthen public safety and security, and improve education for youth. The majority of participants were between the ages of 18 to 34 years old. Recent studies have shown

that young Mexican adults are less civically engaged and less likely to vote [43, 44]. Platforms like México Participa offer one opportunity to engage this age group, gather feedback on their needs and priorities and encourage voter turnout.

Representativeness is a significant challenge. It is important to ensure that decision makers do not see an overly biased dataset. In future work, we are exploring statistical methodologies to compensate for such biases as well as implementing targeted campaigns in rural areas and among underserved participants. Another interesting insight from the Mexico case study is that peer-to-peer ratings afford some level of language independence. Participants read each other's ideas in Spanish and provided ratings. Our team in the United States, without Spanish expertise, was able to extract the most insightful ideas using the peer-to-peer ratings from the participants.

VI. CONCLUSION

The design and assessment of development initiatives is increasingly participatory, where decision makers consider feedback from affected populations. Consequently, a new set of digital data collection tools is needed for problem exploration, where researchers collect a mix of quantitative and qualitative data and rapidly analyze the data to reveal relevant insights from data in the field.

DevCAFE 1.0 proposes a peer-to-peer rating model to address qualitative data filtering and a dimensionality reduction model for rapidly analyzing quantitative data. The tool is customizable and mobile-optimized to directly engage communities from even remote developing regions to collectively assess conditions, needs and outcomes of development initiatives. We evaluated it in pilot projects in Uganda and Mexico. Our results suggest our factor analysis with PCA extracts insightful correlations between demographics and a set of quantitative questions, which can be used to inform changes in development initiatives. In the Mexico case study, we found the peer-to-peer rating system engaged participants, 14% of whom rated more than 10 others. These promising results show that DevCAFE not only provides researchers with scalable qualitative/quantitative analysis, but also engages its participants, which is the ultimate goal of participatory methodologies.

Through the case studies, we learned a number of important lessons that inform our future work. In the Uganda case study, literacy and language barriers were significant. We are developing visual- and voice-based interfaces for DevCAFE. Rather than suggesting a textual comment, participants will record their response. We are also exploring new algorithms for collaborative filtering and new statistical methodologies to mitigate some of the biases inherent to such systems.

ACKNOWLEDGMENT

We thank Heather Lofthouse and Temina Madon for their feedback and guidance on the development of DevCAFE.

REFERENCES

- [1] Ken Banks and Erik Hersman. Frontlinesms and ushahidi-a demo. In *Information and Communication Technologies and Development (ICTD), 2009 International Conference on*, pages 484–484. IEEE, 2009.
- [2] Inka Barnett and Jose V Gallegos. Using mobile phones for nutrition surveillance: A review of evidence. 2013.
- [3] Olivier Barnich and Marc Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *Image Processing, IEEE Transactions on*, 20(6):1709–1724, 2011.
- [4] Michael W Berry and Malu Castellanos. Survey of text mining. *Computing Reviews*, 45(9):548, 2004.

- [5] Benjamin Birnbaum, Brian DeRenzi, Abraham D Flaxman, and Neal Lesh. Automated quality control for mobile data collection. In *Proceedings of the 2nd ACM Symposium on Computing for Development*, page 1. ACM, 2012.
- [6] Lawrence D Brown, T Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, pages 101–117, 2001.
- [7] Waylon Brunette, Rita Sodt, Rohit Chaudhri, Mayank Goel, Michael Falcone, Jaylen Van Orden, and Gaetano Borriello. Open data kit sensors: a sensor integration framework for android at the application level. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 351–364. ACM, 2012.
- [8] Waylon Brunette, Mitchell Sundt, Nicola Dell, Rohit Chaudhri, Nathan Breit, and Gaetano Borriello. Open data kit 2.0: expanding and refining information services for developing regions. In *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, page 10. ACM, 2013.
- [9] Kuang Chen, Akshay Kannan, Yoriyasu Yano, Joseph M Hellerstein, and Tapan S Parikh. Shreddr: pipelined paper digitization for low-resource organizations. In *Proceedings of the 2nd ACM Symposium on Computing for Development*, page 3. ACM, 2012.
- [10] Xi Chen, Paul N Bennett, Kevyn Collins-Thompson, and Eric Horvitz. Pairwise ranking aggregation in a crowdsourced setting. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 193–202. ACM, 2013.
- [11] Nilesh Dalvi, Anirban Dasgupta, Ravi Kumar, and Vibhor Rastogi. Aggregating crowdsourced binary ratings. In *Proceedings of the 22nd international conference on World Wide Web*, pages 285–294. International World Wide Web Conferences Steering Committee, 2013.
- [12] Scott C. Deerwester, Susan T Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by latent semantic analysis. *JASIS*, 41(6):391–407, 1990.
- [13] Nicola Dell, Nathan Breit, Timóteo Chaluco, Jessica Crawford, and Gaetano Borriello. Digitizing paper forms with mobile imaging technologies. In *Proceedings of the 2nd ACM Symposium on Computing for Development*, page 2. ACM, 2012.
- [14] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. Gamification. using game-design elements in non-gaming contexts. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 2425–2428. ACM, 2011.
- [15] John M Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440, 1990.
- [16] Marisol Estrella and John Gaventa. Who counts reality?: Participatory monitoring and evaluation: a literature review. Institute of Development Studies Brighton, 1998.
- [17] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1175–1184. ACM, 2010.
- [18] Anne Marie Goetz and Rob Jenkins. Hybrid forms of accountability: citizen engagement in institutions of public-sector oversight in India. *Public Management Review*, 3(3):363–383, 2001.
- [19] Phillip Good. *Permutation tests*. Springer, 2000.
- [20] Irene Guijt, Mae Arevalo, and Kiko Saladores. Participatory monitoring and evaluation. *PLA Notes 31: Participatory Monitoring and Evaluation*, page 28, 1998.
- [21] Carl Hartung, Adam Lerer, Yaw Anokwa, Clint Tseng, Waylon Brunette, and Gaetano Borriello. Open data kit: Tools to build information services for developing regions. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, page 18. ACM, 2010.
- [22] Marc Helbling and Anke Tresch. Measuring party positions and issue salience from media coverage: Discussing and cross-validating new indicators. *Electoral Studies*, 30(1):174–183, 2011.

- [23] GOLD Judy, Harriet ANDREWS, Gabrielle APPLEFORD, Boni RAMANANTSOA, Odile HANTRINAINA, Rostom DEIPARINE, and Nick CORBY. Using mobile phone text messages (sms) to collect health service data: Lessons from social franchises in kenya, madagascar and the philippines. *Journal of Health Informatics in Developing Countries*, 6(2), 2012.
- [24] JH Jung, Christoph Schneider, and Joseph Valacich. Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science*, 56(4):724–742, 2010.
- [25] Peter Kisaakye. Determinants of unmet need for contraception to space and limit births among various groups of currently married women in uganda. *European Scientific Journal*, 9(19), 2013.
- [26] Jody Zall Kusek and Ray C Rist. Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners. World Bank Publications, 2004.
- [27] Brandie L. Martin and Eric Abbott. Mobile phones and rural livelihoods: Diffusion, uses, and perceived impacts among farmers in rural uganda. *Information Technologies & International Development*, 7(4):pp–17, 2011.
- [28] Fidele Ngabo, Judith Nguimfack, Friday Nwaigwe, Catherine Mugeni, Denis Muhoza, David R Wilson, John Kalach, Richard Gakuba, Corrine Karema, and Agnes Binagwaho. Designing and implementing an innovative sms-based alert system (rapidsms-mch) to monitor pregnancy and reduce maternal and child deaths in rwanda. *The Pan African Medical Journal*, 13, 2012.
- [29] Sukhada Palkar and Emma Brunskill. Analysis of the impact of errors made during health data collection using mobile phones: exploring error modeling and automatic diagnosis. In *Proceedings of the 3rd ACM Symposium on Computing for Development*, page 24. ACM, 2013.
- [30] Somani Patnaik, Emma Brunskill, and William Thies. Evaluating the accuracy of data collection on mobile phones: A study of forms, sms, and voice. In *Information and Communication Technologies and Development (ICTD)*, 2009 International Conference on, pages 74–84. IEEE, 2009.
- [31] Samuel Paul. Making voice work: the report card on Bangalore's public services, volume 1921. World Bank Publications, 1998.
- [32] Raghu Ramakrishnan, Johannes Gehrke, and Johannes Gehrke. Database management systems, volume 3. McGraw-Hill New York, 2003.
- [33] Al M Rashid, Kimberly Ling, Regina D Tassone, Paul Resnick, Robert Kraut, and John Riedl. Motivating participation by displaying the value of contribution. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 955–958. ACM, 2006.
- [34] Adikeshavalu Ravindra. An assessment of the impact of bangalore citizen report cards on the performance of public agencies. *Evaluation Capacity Development Working Paper*, 12, 2004.
- [35] Howard Rheingold. Using participatory media and public voice to encourage civic engagement. *Civic life online: Learning how digital media can engage youth*, pages 97–118, 2008.
- [36] John Rice. *Mathematical statistics and data analysis*. Cengage Learning, 2006.
- [37] Ognjen Scekcic, Hong-Linh Truong, and Schahram Dustdar. Incentives and rewarding in social computing. *Communications of the ACM*, 56(6):72–82, 2013.
- [38] KateWellard, Jenny Rafanomezana, Mahara Nyirenda, Misaki Okotel, and Vincent Subbey. A review of community extension approaches to innovation for improved livelihoods in ghana, uganda and malawi. *The Journal of Agricultural Education and Extension*, 19(1):21–35, 2013.
- [39] William Foote Ed Whyte. *Participatory action research*. 1991.
- [40] Gebremariam Woldemicael and Roderic Beaujot. Currently married women with an unmet need for contraception in eritrea: Profile and determinants. *Canadian Studies in Population*, 38(1 2):61–81, 1944.
- [41] Jun Zhang, Yong Yan, and Martin Lades. Face recognition: eigenface, elastic matching, and neural nets. *Proceedings of the IEEE*, 85(9):1423–1435, 1997
- [42] Center for Sustainable Rural Livelihoods, Iowa State University. Improved nutrition, stable incomes, and hope for the future: Donor impact report. 2014
- [43] Instituto Mexicano de le Juventud, IMJUVE. Encuesta Nacional de Valores Juventud, 2012. Retrieved from http://www.imjuventud.gob.mx/imgs/uploads/ENVAJ_2012.pdf
- [44] Instituto Nacional Electoral, INE. Informe País sobre la calidad de la ciudadanía en México, 2014. Retrieved from: http://www.ine.mx/archivos2/s/DECEYEC/EducacionCivica/Informe_pais_calidad_ciudadania_IFE_FINAL.pdf
- [45] UNHR Participatory Assessment Guidelines <http://www.unhcr.org/450e92a72.html>
- [46] United Nations Food and Agriculture Office Guidelines [http://www.fao.org/docrep/x5307e/x5307e04.htm#chapter two: participatory assessment](http://www.fao.org/docrep/x5307e/x5307e04.htm#chapter%20two%3Aparticipatory%20assessment)
- [47] Dongier, Philippe, Julie Van Domelen, Elinor Ostrom, Andrea Ryan, Wendy Wakeman, Anthony Bebbington, Sabina Alkire, Talib Esmail, and Margatet Polski. "Community driven development." World Bank Poverty Reduction Strategy Paper (2003).
- [48] Prokopy, Linda Stalker. "The relationship between participation and project outcomes: Evidence from rural water supply projects in India." *World development* 33, no. 11 (2005): 1801-1819.
- [49] Dasgupta, A., & Beard, V. A. (2007). Community driven development, collective action and elite capture in Indonesia. *Development and Change*, 38(2), 229-249.
- [50] Casey, Katherine, Rachel Glennerster, and Edward Miguel. Reshaping institutions: Evidence on aid impacts using a pre-analysis plan. No. w17012. National Bureau of Economic Research, 2011. http://www.inegi.org.mx/est/lista_cubos/consulta.aspx?p=pob&c=1
- [51] Nelimarkka, Matti, Brandie Nonnecke, Sanjay Krishnan, Tanja Aitamurto, Daniel Catterson, Camille Crittenden, Chris Garland et al. "IPP2014: Crowdsourcing for Politics and Policy Matti Nelimarkka: Comparing Three Online Civic Engagement Platforms using the "Spectrum of Public Participation" Framework." *Policy* 2012 (2014): 2010.
- [52] Krishnan, Sanjay, Jay Patel, Michael J. Franklin, and Ken Goldberg. "A methodology for learning, analyzing, and mitigating social influence bias in recommender systems." In *Proceedings of the 8th ACM Conference on Recommender systems*, pp. 137-144. ACM, 2014.
- [53] Mo Zhou, Alison Cliff, Allen Huang, Sanjay Krishnan, Brandie Nonnecke, Kanji Uchino, Samuel Joseph, Armando Fox, Ken Goldberg. M-CAFE: Managing MOOC Student Feedback with Collaborative Filtering. *L@S* 2015: 309-312