Adaptive Management Toolkit for Digital Engagement Mechanisms

Table of Contents

Overview ........................................................................................................................................... 1
Design phase ..................................................................................................................................... 2
  Initial appraisal tool ...................................................................................................................... 2
  Annotated checklist for soliciting, assessing and designing digital engagement mechanisms ................................................................. 7
  Analyzing, visualizing and reporting feedback data: some examples ......................................... 14
  Costing template .......................................................................................................................... 22
Implementation and continuous improvement ............................................................................. 26
  Key performance indicators matrix for digital engagement mechanisms .............................. 26
  User satisfaction survey .............................................................................................................. 27
  Guidelines for experiments on aspects of the feedback system ................................................. 28
Hand-over ....................................................................................................................................... 32
  Checklist for assessing conditions for hand-over ................................................................. 32
Overview

The present toolkit is aimed at assisting World Bank staff and clients that implement digital engagement mechanisms (DEM) to performance manage such mechanisms and introduce adaptations for their continuous improvement.

The development of this toolkit has been informed by the review of the pilot implementations of OnTrack, a specific World Bank led innovation using mobile and internet to collect citizen feedback. The review of OnTrack has pointed to the specific challenges the solutions to which have been included in this adaptive management toolkit. Nevertheless, the toolkit is written to support any digital citizen feedback effort promoted or supported by the World Bank. It is adapted from Keystone’s core methodology, Constituent Voice™.

The toolkit includes a series of tools for the management of digital engagement systems in three important phases:

1. Design
   - An initial appraisal tool is offered to help the teams assess the readiness conditions for implementing a digital engagement mechanism and adapting the implementation strategy accordingly.
   - Most often WB teams and PIUs hire external vendors for the development of such mechanisms. To assist them in this process, we include an annotated checklist for soliciting, assessing and designing digital engagement mechanisms. This can be used as a guide for developing the Terms of Reference for vendors, as well as for assessing the received proposals and supervising the development of the mechanism.
   - In addition, we offer a series of examples for analyzing, visualizing and reporting feedback data, as those specifications should be built in the system from the design stage.
   - A costing template provides the list of items to be considered when developing a budget for a DEM.

2. Implementation and continuous improvement
   - A small set of Key Performance Indicators along with suggested data collection methods are proposed as the basis of a monitoring plan for digital engagement mechanisms.
   - Constituent Voice surveys, in the form of follow up surveys, should be applied to collect feedback on key aspects of the digital engagement mechanism itself. We include guidelines for how to implement these.
   - We also include a simple approach to carrying out experiments (in the form of A/B testing for online platforms, when they exist, and the outreach and awareness raising regarding the mechanism itself), as part of a continuous improvement culture.

3. Hand over
   - Lastly, a checklist is provided for assessing the level of local ownership and institutionalization of the digital engagement mechanism before the WB proceeds to a complete hand over to the client.
Design phase

Initial appraisal tool

Introduction
The choice to introduce a DEM as part of a citizen engagement process is a deliberate one. Digital is not automatically better when it comes to citizen engagement. In other words, the design of a DEM should be based on a systematic analysis of the context into which it seeks to intervene. Unless it is rooted in this kind of detailed understanding, expectations for the effort are likely to be unrealistic and the resulting tools ineffective.

This assessment tool is designed to help the World Bank and its clients to assess the context into which they plan to undertake citizen engagement so that they can adapt their implementation strategy to the context, thereby increasing the likelihood of success. It will help them develop realistic expectations about what is possible, identify and build on existing practice, select appropriate communication options and technologies, identify the key obstacles that stand in the way of success, and develop strategies to address these obstacles.

It helps project teams assess the context at two levels: country level and project/PIU level. The country level context helps identify the broad opportunities and constraints. Project teams might be able to influence some aspects of this context (such as government policies, supportiveness of telecoms companies, etc.) but they have little direct influence over many of the factors and would have to adapt their strategies to the opportunities and constraints.

The country level assessment helps the project team find answers to these questions:

- To what extent does the political and bureaucratic culture of government encourage and enable citizen feedback on government services?
- In what ways and to what extent do citizens, individually and in organizations, actively monitor public service policies and performance and engage with government about their needs and government performance levels?
- To what extent do different social groups have access to digital information technology, and how do they currently use these technologies to communicate with service providers?

The country level assessment should not be used for excluding fragile states from introducing DEMs, but should help adapt the strategy accordingly to the in-country conditions.

At project/PIU level many of the factors fall much more within the project’s sphere of influence and control. Task team leaders (TTLs) can tailor their strategies to the opportunities and constraints that they identify – but they can also identify areas that need strengthening before rolling out a citizen engagement strategy – thus greatly improving the likelihood of success. The country level assessment helps the project team find answers to these questions:

- How strongly are PIUs mandated to listen to affected populations?
- How effectively does senior management (both in the WB and the PIU) encourage and support staff to listen systematically to citizens affected by project operations?
• What human, financial and technical resources does the project dedicate to listening to constituents? Are these sufficient to sustain an effective citizen engagement process?
• How effectively does the PIU communicate with its constituent populations?
• How does the PIU currently collect and use feedback from affected populations?

Assessment and scoring
The tool contains spaces to capture three things:

• Brief notes on the opportunities and constraints presented in the situation.
• A simple scoring system (see below).
• Brief notes on the possible implications of the situation for DE strategy.

Possible rating scale:
5  Strong opportunities for DEMs
4  Moderate opportunities for DEMs
3  Opportunities and constraints for DEMs more or less balance
2  Significant constraints for DEMs – but some can be overcome
1  Severe constraints – few can be overcome
0  Context is hostile – introducing DEMs at this time unlikely to be fruitful
Country-level context appraisal tool

Most of the sources of information on which this assessment is based would come from secondary information sources and/or interviews with key informants, such as members of the CMU.

1. A political culture of responsive engagement in government

<table>
<thead>
<tr>
<th>Question/indicator</th>
<th>Opportunities</th>
<th>Constraints</th>
<th>Score</th>
<th>Implications for DE strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Do national government and departmental policies encourage citizen engagement?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 Do public statements encourage government departments to engage citizens?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 Do laws and departments promote public access to government information?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4 Does the government tolerate public criticism and protest?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5 Do national government service facilities seek and respond to citizen feedback and complaints?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6 Do local and regional government facilities seek and respond to citizen feedback and complaints?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. **A culture of citizen engagement with government and private sector** (Answer for high income/education, middle income/education, low income/education. Consider differences in age, sex and language/culture)

<table>
<thead>
<tr>
<th>Question/indicator</th>
<th>Opportunities</th>
<th>Constraints</th>
<th>Score</th>
<th>Implications for DE strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Are citizens aware of their rights and willing to engage government directly and in the media about its policies and performance?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2 Do citizens engage government openly without fear of persecution?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3 Do citizens and citizen organizations make use of transparency laws to access and publicize information on government performance?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4 Do citizens commonly express their views about their service experience and expect a response?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 Are there local leaders and interest groups that claim to speak for communities and inhibit open engagement? (traditional leaders, male leaders)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.6 Are there credible civil society organizations to engage with that enable marginal groups to express their views?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. **Citizen access to information technology and use patterns** (answer for high income/education, middle income/education, low income/education. Consider differences in age, sex and language/culture)

<table>
<thead>
<tr>
<th>Question/indicator</th>
<th>Opportunities</th>
<th>Constraints</th>
<th>Score</th>
<th>Implications for DE strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>What proportion of the population has access to a mobile phone?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>What are the patterns of mobile phone access in different social groups? (e.g., personal device, shared family device, simple phone, smart phone)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>What are the patterns of mobile phone usage in different social groups? (e.g. personal voice calls, personal text messaging, data services, feedback)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>What proportion of the population has access to the internet?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>What are the patterns of internet access in different social groups? (e.g. personal device, shared family device, smart phone, computer)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>What are the patterns of internet usage in different social groups? (e.g. browsing/information, email, commercial activity)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.7</td>
<td>What proportion of the population is literate in a language that uses Roman script?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Annotated checklist for soliciting, assessing and designing digital engagement mechanisms

Introduction

Effective digital engagement mechanisms operate as a cycle that may be described most conveniently in five steps: Design, Collect, Analyze, Dialogue, Course Correct. A failure to complete at any of these steps breaks the cycle, and virtually guarantees that the mechanism is ineffective. Conversely, the steps are not inherently difficult or complex, and faithful execution of each step will significantly improve the results from whatever service or activity is the subject of the feedback.

The present check list can be used by TTLs and PIUs for developing Terms of Reference for soliciting the services of vendors for the design and implementation of Digital Engagement mechanisms. The checklist also provides a guide for assessing the proposals submitted by potential vendors.

Furthermore, the checklist can be used a road map for undertaking in-house development of a digital engagement mechanism and/or supervising such development.

Terms of reference for digital engagement projects should specify requirements for all the elements in this generic checklist. In addition to describing the required features or activities at each step in the operational cycle, the TOR should give some indication of the quantity or scope required. The checklist provides examples for these parameters, which will vary considerably with context.

This checklist assumes that requisite conditions for digital engagement have been previously assessed. To make this assessment, please refer to the Initial Appraisal Tool included in the previous section.

Checklist Summary

Design

1. Review the initial appraisal results to determine if preconditions are in place for the execution of all five steps in an effective digital engagement mechanism.
2. Digital engagement mechanism designs should provide an explanation of the level of evidentiary rigor being sought, the steps taken to realize it, and a method to review and improve evidence quality over time.
3. Digital engagement mechanism designs should consult with constituents to ensure that it is proposing to solicit feedback on issues that are meaningful and relevant for constituents.

1 For a more detailed discussion of methodology for effective feedback mechanisms, see Constituent Voice: Technical Note 1 (version 1.1) (Keystone Accountability, 2014).
4. Digital engagement mechanism designs should include questions that gauge the importance and relevance of the feedback mechanism and its contribution to constituents' sense of agency and voice.

5. Digital engagement mechanism designs should indicate: (a) the ongoing operating cost requires of the proposed digital engagement mechanism; and (b) how the mechanism will provide evidence to justify those ongoing running costs.

6. Digital engagement mechanism designs should indicate: (a) how they will remain non-burdensome to feedback providers; (b) how the mechanism will generate useful quantitative data; and (c) a small number of key performance indicators for the mechanism itself that will be reported at least monthly.

7. Digital engagement mechanism designs should indicate how they would generate feedback across all four categories of information: (i) importance of the service or activity or intervention or organization about which feedback is being sought; (ii) service quality; (iii) relationship quality; and (iv) outcomes for the respondent from the service or activity or intervention or organization about which feedback is being sought.

Collect
8. Digital engagement mechanisms should incorporate a system of testing the user interface to ensure ease of use and clarity, such as A/B testing.

9. Digital engagement mechanisms need a systematic communications plan to ensure that citizens understand the value of providing feedback.

10. Digital engagement mechanisms need a published policy on information use that protects citizen respondent privacy.

Analyze
11. A digital engagement mechanism should contain a data analysis rubric that includes segmentation, triangulation and benchmarking.

Dialogue
12. A digital engagement mechanism must report back the data collected and other program performance information in ways that enable the co-creation of solutions.

13. Reporting back efforts should cover the spectrum from broadcasting to focus groups and key informant interviews.

14. Task front line staff with micro-instigations to specific questions arising from feedback data.

Course Correct
15. A digital engagement mechanism should serve for course correction based on feedback and making sure that the service users know that the changes that they experience resulted from their feedback.

16. A digital engagement mechanism should use ongoing feedback to assess whether course corrections are working.

17. Supplement ongoing KPI review with an annual independent formative assessment.
Step One: Design

The first step is the most time-consuming as the design of a new digital engagement mechanism needs to consider all aspects of the mechanism to be. In subsequent iterations of the cycle this phase only requires a light review.

All digital engagement mechanisms strike a balance between four central principles: evidentiary rigor, sensitivity to process and culture, cost, and utility.

1  **Evidentiary rigor** – The data produced by the digital engagement mechanism and used to form judgments and make decisions should be accurate and reliable.

   ☑ Requirement: DEM designs should provide an explanation of the level of evidentiary rigor being sought, the steps taken to realize it, and a method to review and improve evidence quality over time.

2  **Sensitivity to process and culture** – Development at its best grows the capacity of people to discover solutions and take control of their lives. It is a living, generative process, and this means that a digital engagement mechanism should measure and nurture those things that give life to an intervention: attitudes, relationships, capabilities, voice and agency. These are the stepping-stones developmental processes use to effect the changes in material conditions. Constituents must feel that they belong and contribute.

   ☑ Tip: DEM designs should consult with constituents to ensure that it is proposing to soliciting feedback on issues that are meaningful and relevant for constituents.

   ☑ Requirement: DEM designs should include questions that gauge the importance and relevance of the feedback mechanism and its contribution to constituents’ sense of agency and voice.

3  **Cost** – Given finite resources, organizations must make difficult choices between money spent on digital engagement activities and money spent on the intervention itself. The value proposition of digital engagement is that for a dollar invested in the mechanism, the organization gains multiples of that dollar in terms of development results. The reason for this is that feedback data can laser in on specific problems, generating ideas for solutions, and signaling the effects of corrective actions. Another advantage is that by providing feedback citizens become more mindful of what is happening and not happening, and to adjust their behaviors to improve results.

   ☑ Requirement: DEM designs should indicate: (a) the ongoing operating cost requires of the proposed digital engagement mechanism; and (b) how the mechanism will provide evidence to justify those ongoing running costs.

4  **Utility** – The primary measure of any feedback mechanism is its utility. Does it lead to improvements? Digital engagement mechanisms are designed to generate data that are useful for the key constituents of an intervention or organization – including the citizens who are meant to enjoy the value being created, frontline staff and leadership of the work that is being reviewed by citizens, the financiers, the wider field in which the work is located,
and society-at-large. Digital engagement mechanisms intend to enhance relationships with constituents through more authentic conversations and accelerate outcome attainment. Digital engagement mechanisms are an essential element in a system for continuous improvement in relationship quality and performance management. Digital engagement mechanisms have a real-time orientation and feature short feedback-reflection-action cycles.

Requirement: DEM designs should indicate: (a) how they will remain non-burdensome to feedback providers; (b) how the mechanism will generate useful quantitative data; and (c) a small number of key performance indicators (KPIs) for the mechanism itself that will be reported at least monthly.

Step Two: Collect

Framing conditions
There are three framing conditions for data collection in a digital engagement mechanism.

First, the user interface needs to be simple and clear. Providing feedback should be accessible and non-burdensome.

Requirement: Digital engagement mechanisms should incorporate a system of testing the user interface to ensure ease of use and clarity. One approach to this is A/B testing, discussed in elsewhere in this toolkit.

The second essential element for data collection is that users need to see the value in providing feedback. Digital engagement mechanisms begin with an explanation of purpose with intended respondents. This explanation typically includes the following elements:

- A statement of the purpose of the mechanism and how it fits into a larger purpose of ongoing dialogue and continuous improvement.
- An outline of and timeline for the steps that will follow from data collection, including reporting back to constituents on the received feedback plus other information material to the role of citizen monitors, and a commitment to keep collecting feedback in order to see if resulting corrective measures are working.²
- Some examples of the kinds of things that the organization/service hopes will happen as a result of the implementation of the engagement mechanism. This helps respondents get a tangible sense of the possible benefits of their participation.

These framing conditions for effective digital engagement mechanisms are repeated continuously as part of a long-term culture building and affirming process.

² A recent World Bank-funded experimental study illustrates the importance of supplemental performance information provision. This study of rural community monitoring found that “[e]fforts to stimulate beneficiary control, coupled with the provision of report cards on staff performance, resulted in significant improvements in health care delivery and health outcomes in both the short and the longer run. Efforts to stimulate beneficiary control without providing information on performance had no impact on quality of care or health outcomes.” (http://econ.worldbank.org/external/default/main?pagePK=64165259&theSitePK=469382&piPK=64165421&menuPK=64166093&entityID=000158349_20140826090814&cid=DEC_PolicyResearchEN_D_INT; retrieved 5 September 2014). In other words, when constituents were given information about organizational performance, the whole community monitoring effort was more effective in generating improvements in outcomes.
Requirement: Digital engagement mechanisms need a systematic communications plan to ensure that citizens understand the value of providing feedback.

The third framing condition concerns citizen privacy rights. The digital engagement mechanism must have a clear statement regarding the information that it holds with respect to those providing feedback. Feedback providers must actively consent to the stated use of their information.

Requirement: Digital engagement mechanisms need a published policy on information use that protects citizen respondent privacy.

Outreach for data collection: intentional and always on

Generally, more feedback from more citizens is better than less. It requires proactive outreach to solicit feedback to maintain the kind of volume that enables sustained improvements in services.

Typically, digital engagement mechanisms are “always on” in the sense that it is possible at any time to provide feedback through the mechanism. The interface may be through a computer’s web browser or mobile phone, but the opportunity to click and give an opinion or review is always available. Moreover, the digital engagement mechanism should determine and provide a level of training and support required by the ordinary citizen to provide feedback effectively.

Tip: To realize satisfactory levels of feedback, it is necessary to proactively solicit feedback through intentional surveys and campaigns to utilize “always on” channels. Training and support may always be required.

Step Three: Analyze

Feedback data from digital engagement mechanisms do not by themselves surface and develop solutions. They often have powerful diagnostic value but in themselves are often not enough to guide a full management response. They signal where there are issues to address, and often provide clues to action. And when you take action, ongoing feedback will tell you if corrective measures, once taken, are working.

The development of effective feedback-informed solutions brings us to the other steps in the operational cycle, “Analysis” and “Dialogue”.

Quantified perceptual data can be analyzed using standard statistical methods to give actionable insights into the perceptions of different groups of constituents, and that by tracking these measures of perceptions and their analyses over time it is possible to refine ever more powerful insights. The emphasis in our analytics is to generate clear conclusions and to represent those findings in simple graphics. Examples for reporting and visualization of feedback data are included in a later section of this annex.

In addition to descriptive statistics, analyses should cover three main types of analysis – segmentation, triangulation and benchmarking.

Segmentation disaggregates feedback in one or both of two ways, by the characteristics of the respondent (e.g., gender, age, income level, etc.) or by a response to specific questions (e.g., how likely to recommend the service to a friend).
Triangulation usually draws from different data sources or using different methods of research (classically, quantitative vs. qualitative). For example, in-depth interviews could be conducted with select groups to gain deeper insights into their perspectives on program outcomes.

It is particularly useful to discover consistent correlations between feedback data from the digital engagement mechanism and later occurring outcomes measured independently (and preferably objectively). The customer satisfaction industry provides the most famous example of this. After fifty years and innumerable studies, no one doubts that customer loyalty properly measured in surveys today is an accurate predictor of profits, shareholder value and corporate growth.

Benchmarking is the term commonly used to compare the feedback of a specific digital engagement mechanism with another similar digital engagement mechanism. Benchmarking deepens your understanding of what a particular numerical answer means by showing it in relation to other scores. By comparing against the average, you understand what normal means. And by comparing against the top outliers, your imagination of what is possible is stretched. Say the manager in Region A sees that his region received a mean average score of 7.6. He might be content with this until he sees that Regions B-G have a mean average score of 8.3. A close look at Regions B-G also reveals that the top scoring Region F has a score of 9.4, including only 2 percent giving scores less than 7. The manager in Region A knows what a top performer can do to lift scores, and can now ask the manager in Region F how she did it.

Requirement: A digital engagement mechanism must develop a data analysis rubric that includes segmentation, triangulation and benchmarking.

Tip: To obtain benchmarks, a digital engagement mechanism can join The Feedback Commons.

Step Four: Dialogue

This is the step where feedback data comes alive, both inside and outside the organization receiving the feedback. Effective digital engagement mechanisms “close the loop” with citizens by reporting back the feedback data collected through the mechanism in ways that enable the co-creation of solutions.

Two things happen when you solicit feedback. One is you get answers. The other is that you raise expectations. By closing the loop with respondents you do two things. You refine and validate the meaning of the answers to the questions. And you manage the expectations created by emerging solutions that are within the actual capabilities of the organization and the respondents.

In the place of unbounded expectations, you now have a two-way dialogic process that grounds constituent expectations in granular data – “This is what we heard you saying. Here is some more information that we collect about our program. This is what we propose to do. What do you think of that? These are some of the factors that constrain our capacity to respond here. Are there other factors that you can bring to a solution here?”
Organizations that move beyond data collection to dialogue not only learn and improve their performance, they underwrite higher participation rates and more frank feedback in the future because constituents clearly see the value in providing feedback.

There are many different ways of reporting back and discussing feedback. The mix of methods employed depends on your objectives. One-way, broadcast and publishing modes – usually on the digital engagement mechanism itself – are often used as first step to get the main findings out there and establish a basis of transparency and accountability. They set the stage for more probing efforts.

Meetings or focus groups can be structured as reporting back on feedback and open enquiries into the questions or hypotheses arising from the data. Some organizations prefer independent facilitation at such report back sessions.

Interviews – by phone or in person – with ‘key informants’ offer another way to explore a set of questions emerging from data collection. Key informants are people who have been identified as being willing and able to deepen on the feedback provided. They can be asked to volunteer for such follow up interviews in the feedback collection stage.

Requirement: A digital engagement mechanism must report back the data collected and other relevant program performance information in ways that enable the co-creation of solutions.

Tip: Reporting back efforts should cover the spectrum from broadcasting to focus groups and key informant interviews.

Informal follow up investigations

The simplest way to discover the answers to questions arising from a digital engagement mechanism is often to ask front line staff to discuss these questions with constituents informally as part of their regular interactions. For example, feedback may show that younger women are like other respondents in most questions, but are consistently less trusting of the organization. This begs the question why. To dig deeper, the organization may launch a time-limited micro-investigation in which front line staff share this finding informally with female and male constituents and keep a record of their possible explanations. These are collected and analyzed to see if a clear pattern emerges as well as a consensus about possible corrective actions.

In addition to generating possible solutions this approach creates value in two ways. Firstly, constituents appreciate the value placed by the organization on their feedback. Secondly, staff members are empowered to become actively curious, to be evidence-based problem solvers for the organizations. Employees who excel at this can be recognized and rewarded.

Tip: Task front line staff with micro-instigations to specific questions arising from feedback data.
Step Five: Course Correct

Use it, or lose it! Citizen engagement is an iterative accountability and performance management process, not an exhaustive research study. DEMs tell you when you have an issue. Feedback data empower managers and especially frontline staff to investigate, experiment and engage constituents to find lasting solutions. Ongoing feedback will tell you whether you have solved the issue.

Development of any kind is a process requiring constant change, and DEMs give you the ability to get back on track when something has gone awry.

Digital engagement mechanisms arm organizations with empirically valid data that may be used to secure buy-in across staff and the users, and create a culture of small, iterative steps to test the way forward. DEMs allow you to build a culture based on improvement rather than blame. This culture based on accountability and transparency is the ultimate guarantor of performance and results over time.

Tip: Think of citizen engagement as an iterative performance management process.

Requirement: Course correct based on feedback and make sure those you serve know that the changes that they experience resulted from their feedback.

Requirement: Use ongoing feedback to assess whether course corrections are working.

Evaluation

The KPIs developed for the digital engagement mechanism (see Key Performance Indicators matrix) should provide a basic health indicator for your digital engagement mechanism. In addition, it is recommended that a bi-annual formative assessment of a digital engagement mechanism be conducted.

Requirement: Supplement ongoing KPI review with a bi-annual independent formative assessment.

Analyzing, visualizing and reporting feedback data: some examples

Feedback data from digital engagement mechanisms do not by themselves surface and develop solutions. Survey data can often have powerful diagnostic value and can illustrate broad patterns in constituent perceptions and experience. Open individual feedback platforms and complaints mechanisms can enable management to respond to a specific local complaint.

If feedback data is going to be used to guide a more systemic management response, it needs to be analyzed, visualized and reported in ways that are clear and compelling to intended data users, from citizens to project managers and government authorities.

The way the data is analyzed and presented must enable all users to quickly see the trends and patterns revealed in the data (what is this data telling us?) – but also guide users to explore and clarify the ‘stories behind the story’ in a structured dialogue about the data (What are the factors that led to this feedback? what can each constituent do differently to improve things? How will we measure these changes going forward?)
Three commonly used forms of data analysis – segmentation, triangulation, and benchmarking – have proven effective in this regard.

**Segmentation**

Feedback data can be usefully disaggregated in two ways to understand patterns in the feedback.

Figure 1 Segmentation example from a client report for a timber company

Firstly, it can be disaggregated to reveal differences in experience and perceptions of pre-identified social groups e.g. by sex, age, geographic location, language group and relative prosperity and social status (often shown by proxy indicators such as education level, source of income etc). This enables management to focus improvement strategies towards particular groups.

Secondly, and purposefully from a performance management perspective, we can illustrate the range of responses to performance rating questions (usually from surveys) by segmenting respondents into

1. The proportion of respondents who are delighted (promoters)
2. The proportion of respondents who are neutral (passive)
3. The proportion of respondents who are dissatisfied (detractors)
This allows for simple and purposeful comparison of facilities in different areas, or comparison of ratings of the same facility over time. Managers and staff can focus strategies on specific groups and also set measurable targets such as ‘next year we must improve our ‘promoter’ score to 70%’.

One of the most widely used metrics in the customer service industry is the Net Promoter Score. The South African Government’s Citizen-based Monitoring initiative is adapting Net Promoter Analysis in its comparative performance reports (illustrated below).

**Triangulation**

When perceptual data is correlated with other data drawn from various sources (typically output and outcome data) interesting insights often emerge. For example, when comparing satisfaction levels with variances in crop production outputs reveals consistent correlations, this can have important explanatory value.

Similarly, complementing feedback survey data with targeted in depth interviews can in turn help explain variances in satisfaction, and inform management responses.

It is particularly useful to discover consistent correlations between feedback data from the digital engagement mechanism and later occurring outcomes measured independently (and preferably objectively). The customer satisfaction industry provides the most famous example of this. After fifty years and innumerable studies, no one doubts that customer loyalty properly measured in surveys today is an accurate predictor of profits, shareholder value and corporate growth tomorrow.
Benchmarking

Benchmarking is the term commonly used to compare the feedback of a specific digital engagement mechanism with another similar digital engagement mechanism. Benchmarking deepens your understanding of what a particular numerical answer means by showing it in relation to other scores. By comparing against the average, you understand what normal means. And by comparing against the top outliers, your imagination of what is possible is stretched.

For example, if the manager in Region A sees that his region received a mean average relationship quality score of 7.6, he or she might be content with this until he or she sees that Regions B-G have a mean average score of 8.3. A close look at Regions B-G also reveals that the top scoring Region F has a score of 9.4, including only 2 percent giving scores less than 7. The manager in Region A knows what a top performer can do to lift scores, and can now ask the manager in Region F how they did it.
The South African government’s citizen-based monitoring of community-level government services that we illustrate briefly below is cultivating “parallel feedback” from citizens in communities as well as frontline staff of the local facilities. Citizens assess the performance of facility management and staff, while staff rate the effectiveness of their own managers as well as the upstream District and Regional offices – which is where many service delivery bottlenecks occur.

Comparative Facility Performance Reports and dashboards will allow for side-by-side comparisons of citizen feedback and that of frontline workers. The patterns in the data, and the dialogues based on the data are proving to be a powerful guide to corrective actions.

Case study: The South African Government’s Citizen-based Monitoring pilot

Background
In August 2013 the South African Cabinet approved a ‘Framework for Strengthening Citizen-Government Partnerships for Monitoring Frontline Service Delivery’. The Department of Planning, Monitoring and Evaluation then initiated a pilot project to design and test a system of Citizen-based Monitoring (CBM) in four government departments including the Department of Health and the South African Police Service. Keystone Accountability was engaged to advise in the design and testing of the system.

The emerging CBM methodology generates rigorous and quantifiable, but practical and cost-effective feedback data from both citizens and frontline staff using cost-effective and scalable data collection technologies. Citizens and staff give their feedback anonymously. Once developed, automated backend processes will then turn the feedback on each facility into clear, easy to understand comparative performance reports that benchmark the facility’s performance ratings against other similar facilities.

These reports are then published in the communities and used to frame transparent citizen dialogues through meetings and local media where solutions are co-created, relationships are strengthened and expectations are managed. This publicly validated performance data can then become part of formal departmental M&E and performance management systems in a way that gives citizens an effective voice and influence in how the facilities are managed. This creates the right incentives for departmental staff – from the top
management down to the frontline staff in public service facilities - to perform their jobs responsively to citizen’s needs and priorities in their contexts.

Although the methodology shares a core set of methodological principles, in each case it is designed and built from the ground up through inclusive multi-stakeholder dialogue and robust piloting in real-world contexts. Examples of graphics from the first pilot facility performance reports are illustrated here.

**An example from the Citizen Performance Report on two police stations**

This report shows the responses of community members to one question from the standard survey conducted in both communities. We show the scores of different police stations together so that each police station can benchmark their scores against others in their cluster or province.

Each police station can also compare its scores to an average score for all the police stations in a cluster or district. In a fully developed CBM system, a police station would be able to compare its results with up to 10 other police stations in similar communities to themselves.

For these questions we use a method for analysing and reporting perceptions that is used by many companies to understand and manage their relationship with their customers. It is called **Net Promoter Analysis (NPA)**. NPA gives us a very simple but powerful way of presenting feedback from citizens and frontline staff on the performance of a health facility.

### To what extent do you agree that police investigate cases seriously and competently?

<table>
<thead>
<tr>
<th></th>
<th>Detractor (%)</th>
<th>Passive (%)</th>
<th>Promoter (%)</th>
<th>Net Promoter Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phuthaditjhaba</td>
<td>60</td>
<td>14</td>
<td>26</td>
<td>-34</td>
</tr>
<tr>
<td>Msinga</td>
<td>82</td>
<td>6</td>
<td>12</td>
<td>-69</td>
</tr>
<tr>
<td>Average</td>
<td>71</td>
<td>10</td>
<td>19</td>
<td>-52</td>
</tr>
</tbody>
</table>

In this example:

- **82% of staff at Police Station A police station, and 60% of staff at Police Station B police station gave scores of 6 or less out of 10. These are the detractors.** They feel that police do not investigate cases seriously or competently.

- **6% of staff at Police Station A and 14% at Police Station B gave a score of 7 or 8. They are reasonably satisfied – but are not enthusiastic.** We call them passives.

- **19% of community members at Police Station A and 26% of those at Police Station B gave a score of 9 or 10.** These respondents are very satisfied with the way police at
this station investigate cases that are reported to them. We call them promoters.

When you compare the results of the two police stations, you see that the level of dissatisfaction at both police stations is relatively high, but the dissatisfaction is slightly higher at Police Station A.

In addition to showing the percentage of promoters, passives and detractors, we also calculate a single Net Promoter Score (NP Score). The Net Promoter Scores at both police stations are negative numbers. This shows the high level of dissatisfaction among community members with the performance of both police stations.

Having a single number score makes it easy to calculate Performance Indexes for a group of questions. It also makes it easy to compare changes in citizen performance ratings over time to track improvement or deterioration in performance.

Finally, it is easy to add the scores together so that a provincial or district manager can see and compare the performance scores of all the police stations in a district or a province and calculate a District or Provincial Average Performance Score.

An example of a frontline staff performance report for two clinics
Frontline staff are a key cog in the service delivery wheel. The survey asks staff of their experience and perceptions of:

- Facilities, equipment, resources and materials
- Leadership and supportive relationships
- Motivating and enabling factors (including incentives)
- A short self-assessment

There is also an open question for any other comments. The customer service industry has demonstrated a direct link between staff well-being and satisfied customers (good performance). Staff satisfaction is likely to be a useful predictive indicator for quality services.

<table>
<thead>
<tr>
<th>9.2 My managers are helpful when I have problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHUTHADITJHABA</td>
</tr>
<tr>
<td>COSH</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>

Detactor    Passive   Promoter
In this example:

- 62% of staff at COSH, and 42% of staff at Phuthaditjhaba Clinic gave scores of 6 or less out of 10. These are the detractors. They feel that their managers are not helpful when they have problems.

- 15% of staff at COSH and 17% at Phuthaditjhaba gave a score of 7 or 8. They are reasonably satisfied – but are not enthusiastic. We call them neutrals.

- 42% of the staff at Phuthaditjhaba and 23% of staff at COSH gave a score of 9 or 10, and are very satisfied with the help they get from their managers. We call them the promoters. Phuthaditjhaba has almost twice as many promoters as COSH.

The Net Promoter Scores at both health facilities are negative. This shows a high level of dissatisfaction among staff at both health facilities. However, the dissatisfaction is clearly higher at COSH.

**How to work with these reports**

These Perception Reports are first discussed at meeting of management and staff. It is important that this meeting is conducted in a constructive spirit of searching together to understand the problems staff face and to identify practical solutions together. The meeting could be held somewhere off-site and furniture should be arranged in a way that is conducive to open discussion.

If desired, an independent facilitator could facilitate the discussions. No one should feel that they have to state their personal views in public. The views of staff are reflected in the report. It is these views that are the subject of the discussion.

At this meeting, each participant should have a copy of the report and be encouraged to write down their thoughts in the space provided. The facilitator should lead the group systematically through the report. For each question the discussion should focus on:

1. What are the most notable features of the scores given by citizens and staff?
2. What could be the reasons why a particular score was given? What could explain the differences or similarities in the scores given?
3. What are the possible actions that can be taken to improve performance?

The Citizen Performance Report together with the management response can then be shared in the community, and an open dialogue convened with the support of citizen organizations. The facility management, staff and citizens can identify clear goals to aim for in the next six months or year.
**Costing template**

The following is a costing template that can be used by WB teams to clearly articulate the financial model of a digital engagement mechanism, including legacy costs, and which can be communicated up front to clients.

**Notes to using the template**

*Notes 1-4 introduce the template. Subsequent notes refer to specific lines.*

1. The purpose of this template is to provide a detailed articulation of the costs involved in three distinct phases of digital engagement projects: design & build, operations and evaluation. The template does not refer to which parties are responsible to meet which costs, which is determined by the phase in question and any arrangements agreed between responsible parties.

2. Template assumes three main parties: World Bank, Client (government), and digital engagement project Implementer. In some circumstances the Client will also be the Implementer.

3. The activity list is meant to be exhaustive. Select only the activities that apply to your case. For details of the work that makes up these activities, see the accompanying "Appraisal Tool" and "Annotated checklist for assessing and designing digital engagement mechanisms".

4. Unit numbers and values are typical and need to be adjusted to the context.

5. Can be a national, international consultant or both. Normally one is necessary for initial assessment, but there may be times when you want a team of each. The international expert could be an internal World Bank specialist.

6. Various average consultant day rates are used to indicate typical rates for the indicated service. Actual rates will vary and should be substituted to create more accurate budgets.

7. Levels of inputs for capacity building and support are likely to vary more than other values in this template and will follow directly from findings from the initial assessment.

8. Assumes technology platform development costs reduce as these mechanisms become more common and the local build requires less original work and is readily customizable from available technology.

9. Operation costs are per year. They only include direct costs to core activities of the digital engagement mechanism. Details for these activities provided in "Annotated checklist for assessing and designing digital engagement mechanisms".

10. These day estimates cover the expected level of staff input to operate a digital engagement mechanism over one year. They correspond to the activities provided in greater detail in the "Annotated checklist for assessing and designing digital engagement mechanisms".
11. Benchmarking is an important way to interpret data, providing clear incentives for improved performance. Benchmarking services, such as the Feedback Commons (www.feedbackcommons.org) are becoming available.

12. As needed, an independent formative evaluation can review the efficiency and performance.

13. As needed, undertake an impact study to understand what is different because of the digital engagement project.
## Template with indicative costs

<table>
<thead>
<tr>
<th>Activity</th>
<th>Unit description</th>
<th>No. of units</th>
<th>Unit value in USD</th>
<th>Subtotal in USD</th>
<th>Total in USD</th>
<th>Note No.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design &amp; Build</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1. Initial assessment - international | | | | | | 5
|   Int'l consultant time | Days | 4 | 750 | 3,000 | 3,000 | 6 |
|   Int'l consultant travel | Air or other | 1 | 1,500 | 1,500 | 1,500 | |
|   Int'l consultant per diem | Food & lodging/day | 3 | 200 | 600 | 5,100 | |
| 2. Initial assessment - local | | | | | | 5
|   Consultant time | Days | 5 | 350 | 1,750 | 1,750 | 6 |
|   Consultant out of pocket | Local travel, lodging | 2 | 300 | 600 | 2,350 | |
| 3. Develop terms of reference for the digital engagement mechanism | | | | | | |
|   Consultant time | Days | 3 | 750 | 2,250 | 2,250 | |
| 4. Procurement of CE project Implementer | | | | | | |
|   Advertisement | Advertisements | 1 | 500 | 500 | 500 | |
| 5. Induction of Client & Implementer | | | | | | |
|   Consultant time | Days | 2 | 350 | 700 | 700 | |
|   Consultant out of pocket | Local travel, lodging | 1 | 300 | 300 | 300 | |
|   Consultant travel (if international) | Air or other | 1 | 1,500 | 1,500 | 1,500 | |
|   Consultant per diem (if international) | Food & lodging/day | 2 | 200 | 400 | 400 | |
|   Workshop | Venue, misc. | 1 | 1,000 | 1,000 | 3,900 | |
| 6. Build and develop digital engagement mechanism | | | | | | 6, 7
<p>|   Capacity building for Client and Implementer | Training | 2 | 500 | 1,000 | 1,000 | |
|   Ongoing technical support for Implementer for one year | | | | | | |</p>
<table>
<thead>
<tr>
<th>No.</th>
<th>Activity</th>
<th>Unit description</th>
<th>No. of units</th>
<th>Unit value in USD</th>
<th>Subtotal in USD</th>
<th>Total in USD</th>
<th>Note No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consultant time</td>
<td>Days</td>
<td>12</td>
<td>350</td>
<td>4,200</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consultations with users during product development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consultant time</td>
<td>Days</td>
<td>6</td>
<td>350</td>
<td>2,100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Software development</td>
<td>Days</td>
<td>60</td>
<td>500</td>
<td>30,000</td>
<td>37,300</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Subtotal</strong></td>
<td></td>
<td></td>
<td><strong>50,900</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operations (per year)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Technology hosting</td>
<td>annual fee</td>
<td>1</td>
<td>300</td>
<td>300</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>8.</td>
<td>Ongoing development of the mechanism (technology) -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>consultant</td>
<td>Days</td>
<td>6</td>
<td>500</td>
<td>3,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Communications - consultant</td>
<td>Days</td>
<td>6</td>
<td>500</td>
<td>3,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Data analysis - consultant</td>
<td>Days</td>
<td>6</td>
<td>350</td>
<td>2,100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Dialogue - consultant</td>
<td>Days</td>
<td>12</td>
<td>350</td>
<td>4,200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Reporting - consultant</td>
<td>Days</td>
<td>6</td>
<td>350</td>
<td>2,100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Review and improvement of the mechanism - consultant</td>
<td>Days</td>
<td>6</td>
<td>350</td>
<td>2,100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Subscription to feedback data benchmarking</td>
<td>annual fee</td>
<td>1</td>
<td>1,000</td>
<td>1,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Subtotal</strong></td>
<td></td>
<td></td>
<td><strong>18,900</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evaluation &amp; learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Formative evaluation</td>
<td>Study</td>
<td>1</td>
<td>20,000</td>
<td>20,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Summative evaluation</td>
<td>Study</td>
<td>1</td>
<td>35,000</td>
<td>35,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Subtotal</strong></td>
<td></td>
<td></td>
<td><strong>55,000</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td><strong>124,800</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Implementation and continuous improvement

**Key performance indicators matrix for Digital Engagement mechanisms**

In the following table we provide a small set of key indicators that can be used to monitor the performance of Digital Engagement mechanisms. For each indicator, we provide explanations as well as suggested data sources.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Explanations on the indicator</th>
<th>Data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of awareness of users about the feedback mechanism</td>
<td>Is the target population aware of the existence of the mechanism, its purpose and how it can be used?</td>
<td>Periodic survey of target population (e.g. inclusion in omnibus survey) and/or Focus group discussions during site visits</td>
</tr>
<tr>
<td>Ease of use of the DEM</td>
<td>As perceived by the end users. This indicator can also be used as a proxy for adaptation of the system to the local context and culture.</td>
<td>Follow up survey (Automatically generated after user submits feedback)</td>
</tr>
<tr>
<td>Inclusiveness of DEM</td>
<td>Analysis of the profile of DE mechanism users in terms of:  - Gender  - Age  - Ethnic background  - Socio-economic situation</td>
<td>Data collected through the feedback submission process and/or Follow up survey and/or Website analytics (where applicable)</td>
</tr>
<tr>
<td>Usage rate</td>
<td>Level of usage of the DE mechanism. Depending on the type of system put in place, it may translate to:  - Number of feedback reports submitted in a certain period (e.g. monthly), through different means (SMS, online)  - Response rate (when feedback is solicited through the form of a survey)  - Number of registered users  - Number of unique website visitors in a certain period</td>
<td>System logs and/or Website analytics</td>
</tr>
<tr>
<td>Timeliness of response</td>
<td>Analysis of the timeliness of response to feedback received by the service</td>
<td>System logs</td>
</tr>
</tbody>
</table>

---

3 In the next section we provide a tool for carrying out these follow up surveys
provider. Triangulation of two pieces of data:
- Time passing from feedback being received to generating an effective response (i.e. providing a solution or a plausible justification -- see “fix rate” below) to the user
- Perception of timeliness of response by the user

<table>
<thead>
<tr>
<th>Fix rate</th>
<th>Percentage of issues reported by users that receive a satisfactory solution. Triangulation of two pieces of data:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Information generated by DEM on issues considered “closed”</td>
</tr>
<tr>
<td></td>
<td>- Level of satisfaction of users with solution/response received</td>
</tr>
</tbody>
</table>

And
Follow up survey
(Automatically generated after response is produced)

**User satisfaction survey**

Digital engagement mechanisms should integrate a continuous micro-survey system that would allow them to collect feedback from the users on the performance of the feedback mechanism itself.

The surveys may be triggered automatically at specific touch points with the user, i.e.:
- When feedback is submitted by a user (e.g. an automatically generated SMS message to be sent back to the user, an automated call, or an online survey appearing on a pop-up window when user submits feedback online)
- When a response is communicated by the service provider to the user.
- For those cases where users report problems with a service, after a certain period of time has passed since a response was sent to the user (time will vary according to the specific service in question).

We recommend the use of questions that lend themselves to Net Promoter Analysis, using scales from 0-10. However, it is important to consider this aspect on a case-by-case basis, as in certain contexts, due to cultural and educational factors, the use of these scales may not be appropriate.

Users should receive only 1 or 2 follow up questions at the time, hence limiting the time and effort required from their part for responding. When there is a financial cost associated with the response (ex. sending an SMS), the DEM should consider the offering of a refund or other incentive (e.g. participation in a draw for a prize).

Suggested follow up survey questions include:

- How easy was it for you to submit your feedback through [name of DEM]? 0=Very difficult /10=Very easy
- Was your query answered quickly? 0= Response was very low /10=Response was timely
• Are you satisfied with the response that you received regarding your query? 0=Not at all /10=Absolutely
• [In cases where a problem was reported] Was the issue that you submitted solved by [service provider]? Yes/Partially/No
• How likely would you be to use again [name of DEM]? 0=Not at all /10=Absolutely
• How likely would you be to recommend using [name of DEM] to a relative/neighbor/friend/colleague? 0=Not at all /10=Absolutely

**Guidelines for experiments on aspects of the feedback system**

In order to ensure the continuous optimization of the digital feedback system, we suggest to undertake ongoing simple experimentation with various aspects of the feedback system, such as A/B testing of the online platform (where one exists as part of the feedback system) and different kinds of communications with the users of the feedback system.

**A/B testing for an online citizen feedback platform (when applicable)**

**What is A/B testing**

A/B testing is a simple way to test changes on a website or platform against the current design and determine which ones produce positive results. It can help validate whether any new design or change is improving usage and engagement. A/B testing also helps take the guesswork out of platform optimization and enables data-backed decisions in order to ensure that every change produces positive results, which can be measured by a variety of factors. Constantly testing and optimizing an online platform can increase citizen engagement, while providing valuable insight about platform visitors, and how they use the site.

**How it works**

An A/B test involves creating two versions of an online platform — an ‘A’ version (the control) and a ‘B’ version (the variation) – then measuring the effect each version has on a chosen metric known as the conversion rate. The conversion rate metric depends on the specific goal of the platform, but could include the numbers signing up to use the site or receive updates, simple ‘click-through’ numbers, or the number of citizens using the platform for giving feedback.

A/B testing needs to be carefully thought out and well planned for it to be effective. The following pointers serve as a starting point of things to consider:

• Test early and test often: Tests should be run as early as possible when considering a new promotional technique. Ideally, the platform should be optimized as soon as possible, to maximise reach.
• Always test simultaneously: Running tests on both the A and B variations at the same time is vital, to prevent skewed results based on timing.
• Listen to the results: It is important to resist the temptation to listen to instincts if the empirical data is telling you different. If there is doubt or disagreement, it is possible to re-test.
• Allow the test to run for sufficient time: Cutting the test off early allows for more room for error. The same can be said for letting it run too long. A time period between a few days and a couple of weeks, depending on your platform traffic should be sufficient (aim for a minimum of a few hundred test results before drawing any conclusions).
• Run site-wide tests where appropriate: If testing a call to action or a headline that appears on more than one page, make sure it is tested on every page it appears.
• Make sure repeat visitors see the same variation: Avoid repeat visitors who saw variation A on their first visit seeing variation B on their next visit. This means including provisions in the coding to show repeat users the same page until the test is complete.

Now that you have planned well, the next key to a successful A/B test is consistency and control. The data has to be as accurate as possible, and that requires careful execution. The guidance above should produce successful A/B comparisons with sound results on which important decisions can be based on.

Factors to test in A/B testing
There are many things that could be tested. Platform elements that could be tested for a feedback mechanism include:

1. Call to Action Text: Test the exact wording for asking users to submit feedback.
2. Call to Action Position: Where the call to action is placed should also be tested (above body of page, beside body of page, below body of page, within body of page, etc.).
3. Call to Action Style: How the call to action is styled includes whether it’s just a text link or a button, the size, and the colors used.
4. If the platform integrates other elements besides collecting feedback from users, e.g. mapping of projects, the placement of these and how they affect the likelihood of users to effectively submit feedback should be tested. For example, if mapping is too prominent, it may divert users from the main purpose of the platform.
5. Copy Length/Style: The length and formatting of the main body of text can make a large difference in how many people actually read it. Tests can play with different formats (lists, lots of headlines or short paragraphs) as well as different copy all together to see what works best.
6. Images: The images (if any) used can also make a large difference. Tests here can include which images work best, how many images are optimal, and how large those images should be.
7. Different Offers: Tests could also cover different offers presented to readers. These tests should include offers that have similar values to prevent skewed results. For example, one group of visitors could be entered into a prize draw while others could be offered something else.

The unit responsible for running the platform (WB or PIU) should make a list of the elements that need testing, and then agree a strategy for testing them. It is important to finalise a plan prior to starting to test, so that everything gets tested. Once a strategy is in place, there are a number of tools that can help conduct the tests.

Tools for A/B Testing
There are a variety of shareware and freemium tools for conducting and monitoring A/B tests, but there are two basic types;

1. A tool (generally a script) that will randomly deliver one version of a page (A) or the other (B) to visitors.
2. A tool to monitor the results for each page (which also keeps track of which page the visitor was shown).
Some providers can offer both tools. Google offer a free-to-use⁴ tool called Google Analytics Content Experiments, however, we are suggesting the World Bank uses a commercial, but affordable tool, Optimizely.

Optimizely measures the number of visitors who see both version A and B. It also measures the number of visitors who completed an action – for example, signed up or gave feedback. Once enough visitors have run through the test, Optimizely indicates whether the results are statistically significant, and which version – A or B – is optimal.

Optimizely offers a number of affordably priced plans, and include a free 30-day trial for all plans. Like most technology providers, Optimizely offers a free blog.

There are also a number of useful A/B testing resources too, including the free Apptimize Blog, which covers many aspects of A/B testing. It includes posts (and occasional webinars) on continuous A/B testing, best practice and the types of A/B testing to consider.

Examples
One recent example of successful A/B testing was done by the Obama campaign team, who conducted about 500 A/B tests in a 20-month period, increasing donation conversions by 49 percent and sign up conversions by 161 percent. The team used Optimizely, among several tools, and included elements such as design, user interaction, copy text and imagery in their tests.

The Optimizely blog includes a number of other success stories, which help to highlight how best to conduct A/B testing.

Using experiments to optimize communications regarding the feedback mechanism
Experiments could be used to test different aspects of communications with the intended users of a feedback mechanism.

A wide variety of tests can be done for identifying the most effective methods or types of messages for communicating with users, by examining the following questions:

- Which got more people to sign up to something?
- Which lead to more citizen feedback and higher response rates?
- Which method or message was more appropriate for different kinds of demographic groups (women, youth, elders, rural, urban, etc.)?

Testing different methods
This can be particularly useful when there is doubt about whether a certain tool is more appropriate or not for a certain population. For example, when seeking to engage with a certain population with unique cultural characteristics, it may be important to test whether SMS or IVR is the most appropriate method for soliciting feedback. A test could be easily done by randomly assigning one method or the other to different villages or area codes where the feedback mechanism is being implemented. Then you can measure whether

⁴ If the platform site generates 10 million or fewer hits per month, then Google Analytics is free. If the site generates more than 10 million hits per month, then Google offer increased limits as part of Google Analytics Premium.
there was a difference in response rates depending on the method and which method led to a higher response rate.

Similarly, experiments could be used for selecting the most effective methods for communicating about the availability and ‘how to use’ of the feedback system to a certain population. For example, you could test using orientation sessions in some communities, while in others only distributing flyers and posters.

Testing different types of messages
Depending on the method used, you may experiment with different aspects of the communication that is addressed to the users, such as:

- Call to Action Text: Test the exact wording of your call to action
- Use of language and dialect, including ‘text speak’
- Personalisation of message (Dear Sir Vs Dear John Doe)
- Use of first line in SMS to grabs people’s attention
- Time at which message is sent
- Day at which message is sent
- Length and number of messages

5 Text speak is abbreviations and slang commonly used with mobile phone text messaging
Hand-over

Checklist for assessing conditions for hand-over

This tool provides a checklist of aspects to be assessed in order to ascertain the likelihood that a digital engagement mechanism will be effectively implemented and sustained by the client on the cessation of World Bank financial and technical assistance.\(^6\)

The items in this checklist should be considered in addition to the information collected and analyzed on the KPIs presented earlier in this toolkit.

This tool will become more useful as it is used to collect data. At this point we have no data about how scores provided using the checklist may correlate to an effective hand over and local institutionalization of a digital engagement mechanism. By using the checklist, recording the data, and then reviewing the status of digital engagement mechanisms after cessation of World Bank assistance, it will be possible to identify which requirements in this tool most correlate with digital engagement mechanism sustainability and effectiveness.

The checklist is meant to be used internally by WB operations staff. The assessment should be informed by data collected during the implementation of the mechanism, combined with additional interviews carried out by WB staff with PIU as well as civil society representatives.

Bank team members may wish to use an early implementation of the checklist to trigger a formative evaluation exercise that could help improve the mechanism in the run up to hand over.

<table>
<thead>
<tr>
<th>Question/Indicator</th>
<th>Yes/Partially/No</th>
<th>Implications for hand-over</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is there a steady flow of citizen feedback? (informed by data collected on the usage rate indicator - see KPIs section)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>When there is disruption in the feedback flow, is there a mechanism in place for assessing the reasons and course correcting accordingly?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the relevant political authority accepted formal responsibility to maintain the digital engagement mechanism?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has there been adequate budgetary provision by the relevant political authority to sustain the digital engagement mechanism for at least three years?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the PIU have in place a clear workflow for treatment of the feedback received?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^6\) In this context, World Bank finance could be funds from a World Bank client loan that have been designated for a digital engagement mechanism. Or they could be separate funds provided directly by the Bank during the establishment of a digital engagement mechanism.
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does the PIU consistently respond to feedback and provides citizens with updates on the status of their reports?</td>
<td></td>
</tr>
<tr>
<td>Does the PIU systematically analyze the feedback received?</td>
<td></td>
</tr>
<tr>
<td>Does the PIU systematically take action on feedback received?</td>
<td></td>
</tr>
<tr>
<td>Does the PIU integrate the analysis of the feedback received in the monitoring systems for the particular service/project and its information management systems?</td>
<td></td>
</tr>
<tr>
<td>Does the PIU have sufficient communication resources and capacity in place to continue to raise awareness about the feedback mechanism?</td>
<td></td>
</tr>
<tr>
<td>Does the PIU demonstrate full capacity to deal (directly or through local vendors) with all technological aspects of the feedback system?</td>
<td></td>
</tr>
<tr>
<td>Is technology well maintained?</td>
<td></td>
</tr>
<tr>
<td>Are all necessary user manuals in place?</td>
<td></td>
</tr>
<tr>
<td>Do civil society actors consider that the feedback mechanism operates in a satisfactory way?</td>
<td></td>
</tr>
</tbody>
</table>