Measuring Political Will

Lessons from Modifying the Policymaker Ratings Method
ORS Impact is a consulting firm that helps clients clarify, measure, and align around their social impact outcomes using evaluation, outcomes-based strategy and planning, theory of change, and MLE planning and implementation. We support philanthropies, nonprofits, and government agencies in their most challenging and complex work with our expertise in advocacy and policy change, networks and coalitions, systems change, and initiatives and strategies.

Authors
Sarah Stachowiak
Sara Afflerback
Melissa Howlett

© 2016

This brief was developed with support from the Center for Evaluation Innovation, a nonprofit effort that is building the field of evaluation, specializing in areas that are challenging to assess, such as advocacy and systems change.
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
</tbody>
</table>
| 3    | Policymaker Ratings in Practice: Three Scenarios  
Scenario 1: Developing Unlikely Allies for Federal Farm Bill Advocacy  
Scenario 2: Influencing State Policy with Strategic Research and Engagement  
Scenario 3: Utilizing Federal Advocacy to Support the Development of an Education Movement |
| 17   | Cross-cutting Lessons: Method, Process, and Products  
The Method  
The Process  
The Products |
| 27   | Conclusion |
In 2009, Julia Coffman and Ehren Reed published *Unique Methods in Advocacy Evaluation*, describing four new methods for collecting data relevant to advocacy and policy change efforts. These methods sought to be relevant, timely, and efficient—useful approaches for those working in the often fast-paced and dynamic environments of policy change.

One of the methods included an approach for measuring political will. Political will, an often important outcome for advocacy, has been described as the combination of three factors within elected decision makers: opinion about a particular issue plus intensity of that opinion plus the degree of salience, or importance, of an issue. Advocates often seek to move political will as a key way to drive their issue through the policy process; not surprisingly, it can be difficult to easily, accurately and efficiently measure these facets of decision makers’ beliefs. The Policymaker Ratings method was created to do just that, capitalizing on advocates’ existing knowledge and processes for strategizing around advocacy work.

Developed by the Harvard Family Research Project to evaluate a state-level policy change campaign, the method involves having evaluators facilitate a rating process with a group of advocates to gauge three items for each decision maker in a decision-making body.


- **Policymaker level of support**: Individual policymaker support for an issue based on his or her public behaviors or actions on behalf of the issue.

- **Policymaker level of influence**: Policymaker influence on the policy issue based on criteria that research shows relate to policymaker influence.

- **Rater level of confidence**: Confidence in the accuracy of ratings on the first two scales.

The first of the three ratings is most important for understanding political will; the second and third help provide data to interpret the first. Assessing a whole decision-making body helps advocates and evaluators understand what was needed for a "win" and where opposition or lack of support was concentrated. And the method is repeatable, to capture changes in political will over time.

Seeing the promise of this approach, ORS Impact applied it in 2010 with a funder collaborative seeking state-level education reform in Oregon. Happily, we and our client experienced the expected benefits—the client gained useful new insights by engaging in the ratings process with close partners who had different kinds of knowledge about influence and support, while we yielded useful evaluative data. Methodological gold!

In the past five years, we have continued to recommend and apply this advocacy evaluation method with our clients to measure political will. These clients’ advocacy efforts have spanned work at federal and state levels, have included targeted policy changes and broader sets of policy goals, and have ranged from traditional advocacy organizations to grassroots organizers to think tanks.

Given these varied contexts, at times we modified the method to better fit the work at hand. Some modifications have utterly failed, and some hold promise. Through our experimentation, we’ve learned a lot about facilitating the ratings process and sharing results.

In this piece, we share three scenarios in which we applied the Policymaker Ratings methodology, including what issues emerged, what modifications we attempted, and what we learned that might inform when and how evaluators use this method going forward.

Following the scenarios, we lift up our cross-cutting lessons and observations about key aspects of the methodology. Additionally, we document what we have learned about the ratings process and related products. By sharing our array of experiences, we hope to save others from missteps and help build the effective use of this method.

---

To share our lessons, we thought it helpful to ground them in real-life examples. The three case scenarios differ in status—one has been completed, one is currently in process, and one has been developed but is still untested—yet we believe they each illustrate useful things to be shared with the field.

For each scenario, we provide the following:

- **Context** of the advocacy work
- **Modifications** to the rating aspects
- **Takeaways** of the strengths and limitations we observed

In our description of modifications to the method, we review four main aspects of the ratings:

1. **Rating Focus.** What did the organization rate a group of policymakers against?
2. **Sample.** How was the group of policymakers (or other policy-related decision makers) defined and selected?
3. **Rating Scale.** The original method describes four possible ratings of “support” ranging from “not at all supportive” to “extremely supportive.”
How did we modify ratings of support or other rating scales? To what degree did we use ratings associated with influence and confidence?

4. Rating Process. How did we facilitate the rating process among the selected members of the rating team?

We hope this level of detail across the case scenarios will give texture and nuance to our cross-cutting lessons.
Scenario 1:
Developing Unlikely Allies for Federal Farm Bill Advocacy

Context

Farmers and ranchers aren't usually people's first image of conservationists. But from 2010 through 2014, a communications and advocacy group, in partnership with a firm of strategic political consultants, sought to engage mainstream farmers and ranchers to develop—and then advocate for—regional conservation issues. This effort took place during the reauthorization of the U.S. Farm Bill, an omnibus piece of policy within which policymakers would decide in what ways and with how many resources conservation issues could be addressed. These engaged farmers worked with their policymakers as well as administrators in the Natural Resources Conservation Service (NRCS) to see elements in the Farm Bill that would benefit their interests and—ultimately—yield environmental improvements.

We recommended the use of the Policymaker Ratings method as a piece of the overall evaluation as a way to easily capture baseline and follow-up information on members of the Senate Agriculture Committee. While the organizing effort was too targeted to create noticeable movement in the Farm Bill writ large, there was an expectation that some movement in attitudes and levels of support among targeted decision makers would be seen.

Concerns with Original Method: The political consultants, the key holders of political information, were unwilling to rate an entire body of decision makers, such as the full Senate Agriculture Committee. They had concern that the targeted nature of the effort and the low likelihood of seeing noticeable change at the committee level would lead to findings that might falsely suggest failure of the model and agreed only to a modified approach.

 Modifications

Rating Focus: The intended goals for the effort had to do with both how much money was allocated to conservation work and how funding was assigned; yet, all parties agreed this was unlikely to happen in the timeframe and scope of this effort. Given there was not a single policy goal to measure support against, the consultants rated the identified Senators on three aspects of support:
1. Level of support for targeting conservation funding for resource-specific state area focus
2. Prioritizing conservation funding/programs in the Farm Bill reauthorization
3. Level of support for NRCS having authority for targeted conservation funding

Sample: The consultants wanted to only rate the four Senators on the Senate Agriculture Committee who had jurisdiction over the three regions included in the initiative. Their justification—and pushback against a broader approach—had to do with focus and time. Because the funded activities targeted three strategic regions, they did not seek change beyond the four Senators and felt their time would be ill-spent in exploring the beliefs and stances of the other 20 members of the committee. In addition, the presumed timeline for the work was short: the initiative launched in 2010 with the assumption that the Senate Agriculture Committee would hold hearings on the Federal Farm Bill in 2011, mark up legislation in early 2012, and pass legislation on the Senate floor during the summer of 2012. Given the scope and relative brevity of the work, taking the time to do broader ratings was a hard sell. While we had concerns about such a small sample size, we decided to give it a try.

Rating Scale: With input from the political consultants, we modified the scale for “support.” Targeted policymakers were rated as:

- **Against:** Known to have made statements or had prior behaviors demonstrating a counter-stance on the issue
- **Leaning Against:** Known to have made statements that indicated likelihood of support against the desired issue
- **Neutral:** Known to have made statements indicating a position neither for nor against the issue
- **Supportive:** Known to have made statements in support of the issue
- **Strongly Supportive:** Known to have made statements, taken positive actions, and/or encouraged others to take positive action around the issue

The rating team believed this more accurately captured the levels of support that might be observed, namely in that it accounted for the meaningful difference between no support and direct opposition.

This client did not rate policymakers on “influence” or “confidence.” Because they were targeting ratings to those they knew well and who were believed to be influential to the support desired, there was little enthusiasm for additional ratings.

---

5 We began the facilitated process with the original scales from the Policymaker Ratings method. In discussion with the consultants we made modifications and agreed upon criteria for each point on the scale. Unfortunately, the documentation for these criteria were misplaced. These criteria are a re-creation.
Rating Process: The political consultants arrived at the ratings through a facilitated call with ORS Impact at two time points—near the beginning of our engagement and when the evaluation was completed. In addition to rating the targeted policymakers on the three aspects of support, the political consultants rated both where they thought the individuals were at the rating time and where they hoped they would be in one year.

Takeaways

The modifications around the areas of support, the use of current and hoped for status, and the refinements to the scale to account for more adversarial positions provided useful and nuanced data about support.

The small sample size resulted in a number of issues:

- The extension of the timeline for the Farm Bill aggravated the challenges of using a small sample. The project crossed an election cycle during which one of the targeted Senators lost their seat. It also meant new, un-rated individuals became higher priority targets for advocacy efforts. This limited our ability to see meaningful change over time.
- Little could be reported or shared back without risking confidentiality.
- Focusing on a small, well-known subset limited the strategic value of the process, as little new insight was gained through its implementation.

Near the end of the project, we learned that our insistence on a narrow legislative sample missed meaningful changes that occurred through their work at the administrative level. Because of the political context, the initiative shifted from a focus on federal legislators to regional and federal staff. Since they were excluded from the ratings process initially, it limited our ability to show change in that kind of political support over time.  

While it is unlikely that this client would have agreed to such an approach, it could be useful for others to consider forecasting or scenario planning tools as a way to identify the set of actors who may be pursued in an advocacy effort. A good resource for these kinds of planning tools can be found at: http://sparkpolicy.com/tools/overview/exploring-assumptions-context-forecasting/
Scenario 2: Influencing State Policy with Strategic Research and Engagement  
(In Progress)

Context

In 2014, a think tank (the Center) that seeks to bridge the gap between rigorous scientific understanding and complex environmental policy issues engaged ORS Impact to set them up to regularly measure progress toward their policy goals. The Center disseminates timely, credible, and actionable information to decision makers (primarily at the state level, but also in local and federal decision-making bodies) with the intention of informing environmental policies. Though not an advocacy group, they do seek to improve policy decisions in their state by dispelling myths and increasing the use of facts.

Rather than waiting long-term to assess whether sounder environmental policies were being enacted, ORS Impact recommended the Center include the Policymaker Ratings method in its ongoing Measurement, Learning, and Evaluation (MLE). This allows them to conduct regular “pulse checks” on where decision makers stand on priority policy issues.

Concerns with Original Method: Original concerns had to do with the use of an advocacy tool at all. As a nonpartisan think tank, the researchers don’t advocate for a particular policy; rather, they produce research to inform policy decisions. Modifications were necessary to see how political support relative to facts and science-based policy solutions could be tracked over time.

Modifications

Rating Focus: In its strategic plan, the Center had identified two priority issue areas that the majority of their research products and outreach would address in the near-term. Because the prioritized issue areas were broad, the rating team found it difficult to assign decision makers a single rating. Given that challenge, we identified a number of key messages on which to rate decision makers within each priority issue area (because, for example, a decision maker could be supportive on one message but unsupportive of another message within the issue). With two focus areas with six messages each, the method of rating by message grew tedious, and so the Center ultimately chose to carry out the Policymaker Ratings method for messages tied to major publication and outreach efforts. Prior to the
release of a major publication, the team would identify up to four key messages from their research findings to rate a set of decision makers against.

**Sample:** For each publication that is prioritized, the team identifies a unique set of key targets who are then rated. This means that each rating process has a set of decision makers, including legislative and administrative organizations and individuals at the local, regional, state-wide, and federal levels.

**Rating Scale:** Since rating by key message added to the complexity of the process, the team decided to narrow to one scale. We modified the scale to map to the Center’s Theory of Change by including understanding and a continuum of championing:

- **Lack of Awareness or Understanding:** The decision maker does not demonstrate an understanding of the policy issue and/or proposal.
- **Understands; Opposes:** The decision maker demonstrates an understanding of, but disagrees with, the policy issue and/or proposal.
- **Understands; Agrees:** The decision maker demonstrates an understanding of, and agreement with, the policy issue and/or proposal.
- **Understands; Promotes:** The decision maker demonstrates an understanding of, and promotes or champions, the policy issue and/or proposal.

The original rating scale did not delineate between lack of understanding and opposition. Given its work, the Center saw value in assessing understanding as a key step in the process. Additionally, the Center saw instances where a decision maker clearly understood a message, yet advocated against that issue/proposal.

The Center did not rate “influence,” because they perceived the audiences for their publications as influential. The rating team did assign “confidence” ratings at the decision-maker level.

**Rating Process:** At an in-person work session, ORS Impact facilitated the initial few ratings with a small rating team, until they had a sense of how to use the method. The team now rates their targeted list before releasing a publication. Six months post-release of the publication, the rating team rates the same decision makers on the same key messages to assess the degree to which political will shifted as a result of their efforts.
Takeaways

+ The Center had reservations implementing a method that stemmed from advocacy evaluation, given that it does not see itself as a traditional advocacy organization. Instead, they referred to the tool and process as “Position Tracking.” While as a non-advocacy group they had some unique challenges, it still has proven relevant and useful.

+ The rating team found it useful to explicitly identify a few key messages ahead of a major publication release; the intent being, that they could see which messages moved the needle more than others. Additionally, the use of this process has helped hone their strategy around targets for communications.

− It was difficult for the Center to determine which publication and outreach efforts were worth evaluating with the Policymaker Ratings method. Given the volume of research the think tank produces, it is only feasible to go through the rating process for a few publications each year.

− It is not clear how useful “rolling up” data across publications and messages will be for understanding support more broadly. Because each group is its own unit of analysis, they can understand more about their success with particular publications and messages, but not necessarily about how political support is changing more broadly over time. Additionally, having ratings for a mix of individuals and organizations in different domains (e.g., legislative, administrative) in one set of ratings is a challenge for analysis and interpretation, and the process of self-identifying targets could result in selection bias when using the data for evaluative purposes.
Scenario 3: Utilizing Federal Advocacy to Support the Development of an Education Movement (Untested)

Context

In 2014, we began working with an advocacy organization that led a loose coalition of other D.C.-based organizations, all supported by one foundation, to improve federal education policy. While their original approach to measuring their success was to identify annual and multi-year policy change targets, we saw an opportunity for the group to understand and report on their progress over time by tracking policymaker support for the kind of changes in education policy they ultimately sought. We saw this as a ripe opportunity to have a group process around the Policymaker Ratings method that could help this diverse group of organizations share political information and better coordinate their work.

Concerns with Original Method: The client was extremely concerned about confidentiality and had a strong aversion to negatively rating individual policymakers. As a well-known advocacy group, they did not want to put into writing who they viewed as priority decision makers and who they viewed as the opposition, especially when these classifications can change given different political contexts and climates. There was fear that such a document could be released to other organizations or to the public, hurting the group’s reputation. In consultation with the client, we developed an approach that would avoid individually rating policymakers while ultimately providing very similar information and benefits.

The week before facilitating a process to conduct the initial ratings, the Every Student Succeeds Act (ESSA) was signed into law. The implications for federal education policy pivoted the organization’s focus from policy advocacy to policy implementation and, due to the nature of the bill, shifted the focus from federal to state-level work. The transition would introduce a whole new set of actors, and the organization was not ready to pivot to using the process with a new set of key players.

Although we did not end up carrying out the modified process, and cannot speak to its relative success or failure, we still want to share the planned modifications. We hope to find another opportunity to use them in the future; we believe they will be able to address some concerns about using the method.
Proposed Modifications

Rating Focus: The organization planned to rate federal policymakers on their level of support for an education movement the organization is helping to grow nationwide.

Sample: We intended to allow meeting participants to generate a list of key decision makers into categories of varying levels of support (see Rating Scale section). Rather than working through a predetermined list of individuals within the right committees, the group would generate the sample in real time. This emphasis on a “positive” approach allayed some of the concerns about negative individual ratings.

Rating Scale: Given feedback we received during planning conversations with the organization and other experiences we had with federal advocacy efforts, we refined the rating scale as follows:

- **Overall Champion**: Individuals that regularly advocate on behalf of policy that supports education goals aligned with the organization’s strategy
- **Overall Supporter**: Individuals that support education goals aligned with the organization’s strategy
- **Single Issue Supporter**: Individuals that support a single issue (e.g., assessments) though they may not be aligned with all the goals of the organization
- **Potential Supporter**: Individuals that could be cultivated to support aligned education goals, but need additional attention from existing supporters
- **Desirable Supporter**: Individuals that are not yet supporters of the organization’s education goals, but would be valuable as such

The category for single issue supporters, or individuals who support a single issue (e.g., assessments), accounted for those individuals who may not be aligned with the full set of goals for the organization and movement, but could be counted on for certain policy goals. The rating scale allowed for a nuanced analysis of where a decision maker falls on the continuum of “support” but avoided any classifications that could be perceived as negative.

Once the group had generated the names of federal policymakers for each of these categories, we would use a group process to ensure consensus around the placement of different individuals, then identify those with “high” influence. We would also ask them to list other key influencers who had not shown up in any of the existing categories.

We did not plan to rate “confidence” because the group process implies medium to high confidence for all those listed. The assumption was that any decision makers for which the group was unsure of their rating would not make it through the process.
**Rating Process:** ORS Impact would facilitate an in-person meeting with the rating team to conduct the initial ratings. The client would facilitate a similar process with their internal team and partner organizations annually.

Unlike the traditional approach of running through a list of individuals to rate, we would take back the information generated and crosswalk it against lists of key legislative committees. We could then calculate, over time, things like the proportion of identified supporters within meaningful decision-making bodies and the number of influencers who were supporters.

(ANTICIPATED) Takeaways

Although we did not implement this modified method, it still seems useful to raise potential concerns or issues for other evaluators to keep in mind if they experiment with a similar set of modifications.

- We still think there could be promise in this approach, or something like it, for groups who balk at individual ratings. While it puts more onus on the evaluator on the back end, it may be a useful way to engage advocates to reap the evaluative and strategic benefits if they would not otherwise agree to participate in the typical process.

- There is a possibility that inviting the rating team to generate the sample will result in a collection of both positions and entire organizations (e.g., both the Secretary of Education and the National Education Association). It would be difficult to compare across ratings of individuals and organizations. We recommend limiting the rating team to one or the other. Similarly, the generative sample will likely result in both administrative and legislative decision makers and/or bodies. The evaluators and the rating team will need to decide how to treat the two types similarly or differently.

- If the rating team decides to rate individual policymakers (as opposed to decision-making bodies), there is the potential for the sample to result in individuals that do not fit into bodies, individuals that are at vastly different levels, and/or too few individuals within a given body to draw conclusions. While it can still be valuable to generate the sample as part of the initial rating process, the evaluators and the rating team should work together to winnow the sample to like individuals or to apply parameters to the initial sample generated.
## Table 1: Overview of Original Tool and Three Modified Scenarios

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rating Focus</strong></td>
<td>Particular policy proposal</td>
<td>Senators’s support for conservation funding/programs in the U.S. Farm Bill reauthorization</td>
<td>Decision makers’ support for a few key messages from each major publication released</td>
<td>Federal policymakers’ level of support for an education movement</td>
</tr>
<tr>
<td><strong>Sample</strong></td>
<td>All policymakers in a decision-making body (e.g., state legislative houses)</td>
<td>Four Senators on the Senate Agriculture Committee</td>
<td>Primarily state decision makers (legislative and administrative)</td>
<td>List of federal policymakers developed generatively</td>
</tr>
</tbody>
</table>
| **Rating Scale** | Support  
• Not At All Supportive  
• Somewhat Supportive  
• Supportive  
• Extremely Supportive  
Influence (ranging from Not Very to Extremely)  
Confidence (ranging from Not Very to Extremely) | Support  
• Against  
• Leaning Against  
• Neutral  
• Supportive  
• Strongly Supportive | Support  
• Lack of Awareness or Understanding  
• Understands; Opposes  
• Understands; Agrees  
• Understands; Promotes | Support  
• Overall Champion  
• Overall Supporter  
• Single Issue Supporter  
• Potential Supporter  
• Desirable Supporter  
Influence (High only) |
| **Rating Process** | • At least 3-5 advocates participate in the rating process;  
• Process is repeated over time | • Facilitated calls (2) with ORS Impact and political consultants  
• Rated current and aspirational support  
• Measured at the start and conclusion of ORS Impact’s evaluative efforts | • Initially facilitated in-person by ORS Impact with 3-4 staff; process owned by rating team going forward  
• Measured twice, at publication pre-release and after six months | Intended to be facilitated in-person by ORS Impact with 10-12 individuals from 3-4 organizations; to be repeated annually |
| **Takeaways** | When used for a particular policy, process yields useful data for evaluators and advocates | Sample was too small to be useful | Finding the right level of granularity for ratings focus has been a challenge; new approach is useful for informing strategy but hard to “roll up” | For groups reluctant to rate legislators as unsupportive, this may be a useful modification, though it is untested |
Cross-Cutting Lessons:

The Method, The Process, and Products

As illustrated through the scenarios above, the contexts and ways in which organizations seek to move political will can vary widely. Having a method that captures complex advocacy work in a way that is efficient, has strategic benefit, and doesn’t require engagement of the policymakers themselves continues to be attractive to us and our clients. But—as is likely clear by now—there are ways in which this method can be more or less effective. Here we lay out our lessons learned related to the Policymaker Ratings method, including the process for implementing the method, and how you might visualize and share results.
Small samples are not as useful.

We continue to get the most pushback from our clients about the idea of rating an entire decision-making body. Often people feel they have incomplete information or that the “juice” of rating a body is not worth the “squeeze” of doing all the ratings work. However, we find that doing a targeted or small group of ratings—particularly if those ratings are very focused on specific individuals—can be problematic. People change positions. The decision makers that advocates will want to focus on are mutable and sometimes volatile. Given this, push as much as possible for a meaningful body in which you can see change over time. At a minimum, we now focus our clients on looking at support in key positions versus individuals to minimize the impact of turnover, though this activity may be more useful for tracking individual champions over time than understanding political will.

Experimenting with rating scales can be fruitful.

One of the most useful contributions we think we can make is to share some of the ways we have modified and expanded upon the original rating scales. We think expanding ratings to include specific ways to account for both adversarial and neutral positions helps to strengthen the method and provides a more nuanced understanding of political will in many areas. Adding specificity to reduce subjectivity will only continue to strengthen the broader set of rating scales.

There is still a nut to crack for ratings that are not on one specific policy goal.

In our experience, using the Policymaker Ratings method as originally designed, without modification, works extremely well for specific policy goals, like universal pre-K. Yet, political will is important in more contexts than just that. We have struggled with hitting the right rating focus. Too broad (e.g., “early childhood”) and it’s nearly impossible to meaningfully rate individuals (e.g., one legislator is for quality childcare but not for Head Start funding). Too narrow (e.g., a particular advocacy message) and the relevance is often short-lived or the ratings process becomes tedious. We have sought to address this in rating scales by identifying people as “single issue supporters,” but we expect to continue experimenting in this space.
Even when there are evaluative challenges, we always see strategic value.

As shown in our scenarios, some of our modifications have created headaches—the inability to better measure change over time or report back useful findings has frustrated us. Yet in all cases, including those not detailed in this report, we hear back from clients that they do see benefit from going through the process (even when we have to provide a lot of encouragement to do so). Hearing colleagues and partners talk through the policy positions of different decision makers, taking the time to reflect on influence, and even deciding who and how to rate individuals has led to strategic insights that have felt additive and useful for advocates’ work.

There is room to learn more about applying this process with administrative policymakers.

Comprehensive advocacy efforts often include targeted decision makers within various levels of government, including those in administrative positions who have a great deal of influence over the implementation of policies. However, bureaucracies are different from legislative bodies, and the actors within bureaucracies have different motivations and incentives. We’ve found a desire in several cases to assess the general will around particular policy priorities across various relevant bodies, which is not something this tool was designed to do. We will continue testing ways to measure this concept and see if there is an appropriate application of this method to that context.

Confidentiality is key.

The process, data, and results from this exercise are highly sensitive to participants. Given this, ratings should be aggregated before disseminating beyond members of the rating team. It is important to have clear guidelines from the outset about who is involved, who has access to the data, and who will receive aggregated results. With all clients we encourage participants to agree to a “cone of silence” so that the discussion and final ratings stay within the rating team. As far as storing the results, we recommend the organization maintain a password-protected spreadsheet or ask the evaluation partner to store the data on their server.

In any case, there will likely still be trepidation on behalf of participants. Once everyone is in a room together, it can be helpful to address this straight on and open it up to questions or concerns. Ensuring that everyone hears the same questions and answers can help create a space where participants feel they can speak candidly.

Aim for diversity in raters.

Who is included in discussions can greatly impact the success of this exercise. The following can be useful to consider when deciding who to include on your rating team:

- **Size:** It is helpful to have a healthy-sized group that pulls together varied perspectives and stakeholders, including those with the greatest expertise in this area. This may include key staff focusing on programmatic work, advocacy, lobbying, and/or communications from both within the organization and from partner organizations. Perspectives from different roles, different organizations, and different relationships to the decision makers will provide more nuanced ratings than the key staff from the sponsoring organization alone.

- **Positions:** Because rating processes can include groups of people across different roles in an organization, it is important to be aware of possible power dynamics that could mean all participants may not feel they can speak candidly. Assessing dynamics within the rating team beforehand can help prevent, or at least anticipate, potential discomfort during the rating team meeting.
• **Trust:** The exercise is most successful when participants have a sincere interest in working together and feel a level of trust and comfort with one another. To promote trust among members of the rating team, be explicit about who will see and interact with the data, how the ratings will be aggregated, and who the aggregated results will be shared with.

Go for consensus.

The original publication invites participants to either rate policymakers as a group (consensus rating) or independently (average rating). All of the groups we have worked with have used the consensus approach to determine ratings. The practice of arriving at consensus makes the rating process more involved and promotes conversation and learning across members of the rating team. We cannot foresee a situation where the average approach would be a better fit. The hesitation is not with an average rating itself, but with the conversation that would likely be lost with this approach.
There are useful ways to visualize political will or support.

Once the rating team has gone through the rating process, you want to—at a minimum—visualize a “snapshot” of political will (that is, the proportion of policymakers who support or don’t support a policy issue or proposal). As you rate the same sample a second or third time, the data can be visualized to show changes in political will over time (that is, the degree to which policymakers have moved on a policy issue or proposal since they were last rated).

As with other data visualizations, it’s important to keep the audience and purpose at the forefront. Reporting policymaker ratings data to those who have decision-making authority over the advocacy or policy change effort (e.g., boards or leadership) and for evaluative purposes might look different than if you were to visualize the data for those who have a direct responsibility for the advocacy or policy change effort (e.g., staff or grantees/partners) and for strategic purposes. That said, we’ve seen groups use various techniques to visualize changes in political will and have found stacked horizontal bar charts and tree maps to be best suited for telling this type of data story.

*Figure 1: Changes in Policymaker Support (assessed annually 2014-2016)*

<table>
<thead>
<tr>
<th>Year</th>
<th>Strongly Supportive</th>
<th>Supportive</th>
<th>Neutral</th>
<th>Leaning Against</th>
<th>Against</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>7%</td>
<td>41%</td>
<td>14%</td>
<td>28%</td>
<td>10%</td>
</tr>
<tr>
<td>(Baseline)</td>
<td>n = 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>14%</td>
<td>43%</td>
<td>14%</td>
<td>18%</td>
<td>11%</td>
</tr>
<tr>
<td>n = 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>24%</td>
<td>40%</td>
<td>16%</td>
<td>8%</td>
<td>12%</td>
</tr>
<tr>
<td>n = 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Four policymakers in our sample left their positions between the baseline and 2016 ratings.*
The more descriptive data you track, the more you can say.

It may feel tedious to track characteristics of your sample of policymakers (such as governmental and/or legislative body, political party, or geography). Coffman and Reed suggested that this descriptive data “can be added to enable different ways of looking at the analysis.” In our array of experiences, we have found there to be evaluative and strategic benefit in capturing and visualizing policymaker characteristics; it sheds light on where there are disparities in support, and thus can inform future decisions about where to focus your advocacy efforts.

The following are examples of visualizations that these descriptive data lend themselves to.
Political party
(e.g., Republican, Democrat, Independent)
This can also be applied to any other group such as governmental or legislative bodies.

Figure 3: Political Will by Party Affiliation
(n=75)

Democratic (n = 41)
- Extremely Supportive: 38%
- Supportive: 22%
- Somewhat Supportive: 15%
- Not at all Supportive: 15%

Republican (n = 32)
- Extremely Supportive: 50%
- Supportive: 50%
- Not at all Supportive: 15%

Independent (n = 2)
- Supportive: 50%
- Extremely Supportive: 50%

Visualization adapted from Julia Coffman

Geography
(e.g., county, state)

Figure 4: Average Legislative Support Ratings by County
(n=120)

Average Levels of Support
- 1.0 - 1.49 Low Support
- 1.5 - 1.99
- 2.0 - 2.49 Medium Support
- 2.5 - 2.99
- 3.0 - 4.0 High Support

Changes in Support
- Increase from prior year
- Decrease from prior year

Rating Team Confidence
- Not Very Confident
- Somewhat Confident
- Very Confident

0% 30% 70%

Visualization adapted from Julia Coffman
Overlay “influence” and “confidence” data with “support” to show nuance.

Some of the groups we worked with chose not to assess policymaker level of influence. However, we have found strategic value in capturing influence data and visualizing it alongside support data. That is to say, you can showcase these data side-by-side to highlight, for example, the proportion of highly influential decision makers who are not currently supporting the policy issue/proposal so that you can target those individuals in your future advocacy efforts.

The following is an example of policymaker support overlaid with policymaker influence:

*Figure 5: Policymaker Support Relative to Influence (n=24)*

Similarly, data around the rating team’s confidence can be showcased side-by-side with the policymaker ratings, to indicate to your audience whether the results are reliable. See Figures 4 and 5 for examples of how to visualize confidence ratings.
Conclusion

The Policymaker Ratings method was an early innovation in the advocacy and policy evaluation space that continues to have relevance and utility today. We have not yet found another approach for measuring political will that so effectively combines the data benefit for evaluators and the strategic benefit for clients. We set out writing this piece knowing we had lessons, not answers—we hope this will help other evaluators build from our experiences and continue experimenting in this space.