Performance Measure Reality Check[1]

Susan Hostetter and John Walsh
U.S. Census Bureau, Washington, DC, USA

Executive Summary

This paper will provide managers a practical "reality check" process for selecting the most relevant and implementable performance measures for their projects, programs, and organizations. We will cover how to:

- *Evaluate data for relevance, quality, timeliness, and accessibility*
- *Evaluate measures for fit to strategic goals and ease of implementation*

Additionally, the paper will cover tips for organizing your results for discussion with leadership, the benefits of the process and the time and resources required to implement.

[1] *The views expressed in this paper are those of the authors and do not necessarily represent those of the U.S. Census Bureau.*

The Reality Check

Often as a manager you are asked to answer questions about the progress on your project. This is generally a straightforward task, but as you delve into data to support your answers you may find that the reporting process can quickly become overwhelming. In a perfect world you would have time and resources to produce measures of everything and anything to describe your project, but you are in the real world and in the real world you need a reality check.

In our experience we have found that there is always data collected (not directly related to performance measurement), and there are always questions from stakeholders/leadership. Performance Management is the marriage of the two. Typically, you are producing answers for a stakeholder audience and you will have many ideas and may receive many more ideas for measures to track progress for your project. This is where you need a process for identifying the best candidates for development and here is where we can help. In our work with performance measures, we practice a process to systematically review and rate measure choices to reduce the big list to a defendable and socialized list of the best and most feasible choices.

Our performance measure reality check process reviews two sets of criteria: those that describe a measure's purpose and those that quantify the feasibility of producing a measure. The review of purpose criteria is a conceptual process where you will identify insights, questions answered by the measure data and evidence-based actions and evaluate those data points against the goals and purpose of the project. This process will assign values to the measures and elevate those that are perceived to provide the most informative data. The review of the feasibility of producing a measure criterion will quantify your ability to produce each measure. This review evaluates the measures by organizational readiness, data availability, consistency and frequency, respondent and organizational burden and effort to analyze and format the measure. It then assigns values to each of the feasibility criteria to elevate those measures that are ready to be developed. The final step in our process is to combine the purpose and feasibility evaluation results and calculate a final score. The resulting scores then provide values to help you or leadership make decisions on how best to spend project or program resources producing and maturing performance measures.

Demonstration Examples

We included some demonstration examples to help illustrate how to describe and rate the different criteria in our process. These examples are drawn from our work with performance measure development. Table 1 is a list of the measures used.

MEASURES
Percent of Research Proposals Accepted by Customer
Percentage of Change Requests Implemented
Planned Versus Actual Time to Complete Work
Budget Variance
Percent of Staff with Met Target Proficiency Level

Table 1 List of Demonstration Examples

Measure Purpose Evaluation

The process to evaluate the purpose of your potential measures can seem like a superfluous step but with the engagement of leaders and stakeholders in the process will help identify measures that are worthy of the time you will invest to produce them and simultaneously gain leadership and stakeholder buy in of the results. This is an important consideration; measures are an investment. They require time and resources to develop, produce, maintain, and review and you will want those measures to be the right measures. Once you gather the purpose evaluation inputs, you will then use the insights into the program, leadership questions to be answered, and evidence-based actions to assign a value.

Insight into the Program from the Data

We define Insight into the Program from the Data criterion as ways to answer the larger questions about the program and tie program data to the strategy and direction of the organization. Typically, the insights generated for your measures may be more closely tied to the overall strategy and direction of the organization than to the day to day operations of the program being measured. Decisions made because of these insights may be more sweeping and impactful over the long term and not solely focused on short-term progress.

MEASURES	INSIGHT Into the Program
Percent of Research Proposals Accepted by Customer	How desirable/ marketable is DSMD's work to our customer base
Percentage of Change Requests Implemented	The validity and scope of the change requests
Planned Versus Actual Time to Complete Work	Quality of the work planning, Accuracy of the work effort estimates
Budget Variance	Quality of the work planning, Accuracy of the budget estimates
Percent of Staff with Met Target Proficiency Level	Areas where we have needed proficiency, Efficacy of hiring and training programs

Table 2 Examples of Program Insight Responses

Leadership Questions to Be Answered by the Data

As opposed to program insights, we define Leadership Questions to Be Answered by the Data criterion as the more operational, day to day questions that deal with the general health and execution of a given project or program. These questions (and answers) can help to guide the

small changes to keep a project executing at full potential. These questions may be typically tied to the triple constraints or could be focused on other facets of project execution that are more in line with stakeholder interests.

MEASURES	Leadership QUESTIONS to be Answered
Percent of Research Proposals Accepted by Customer	Are the proposals well defined? Do the proposals align with the actual needs of the customers?
Percentage of Change Requests Implemented	How are these requests impacting the overall level of service provided? Is the baselined scope of work incorrect?
Planned Versus Actual Time to Complete Work	Do the planning models work? Is the level of effort being adequately captured?
Budget Variance	Do the planning models work? Are all aspects of the project being adequately captured?
Percent of Staff with Met Target Proficiency Level	Where have we met our proficiency targets? Where do we need to concentrate our efforts?

Table 3 Examples of Leadership Question Responses

Evidence-Based Actions Provided by the Data

We define Evidence-Based Actions Provided by the Data criterion as the potential changes to the operations or strategy that are informed by the answers to the leadership questions and the program insights. This generation of measures that enhance evidence-based actions will support the strategic use of resources and highlight those programs and projects that are working for the organization's strategy and goals and spotlight those that are not.

MEASURES	Evidence-Based ACTIONS
Percent of Research Proposals Accepted by Customer	Evaluate training/skills we need for providing the customers real value, tailor our proposals to feedback and external factors
Percentage of Change Requests Implemented	Review whether the change requests of a given project meet or exceed expectations, pay closer attention to project or management processes where indicated
Planned Versus Actual Time to Complete Work	Consider adjustments to the resource and budget plan, Review change requests to cover unplanned work
Budget Variance	Adjust the budget plan, Review change requests to cover unplanned work
Percent of Staff with Met Target Proficiency Level	Identify needs for hiring, training and knowledge sharing, Adjust training initiatives to new priorities

Table 4 Examples of Evidence-Based Decision Responses

Scoring of Measure Purpose Criteria

The second step of the measure purpose evaluation process is to assign values to the details you captured about each criterion. For determining the values, we focused on the "priority" of the measure as a whole and Table 5 below shows the values in the scale. We have found that a scale of three values: high, medium, and low, is enough to produce a spread of values that is meaningful in the final scoring process. Having more "middle values" only adds confusion to the

scoring process. Table 6 below shows the values we have assigned to each of the demonstration examples.

Value	Priority Level
1	Low
2	Medium
3	High

Table 5 Measure Purpose Scoring Definitions

MEASURES	Score
Percent of Research Proposals Accepted by Customer	3
Percentage of Change Requests Implemented	1
Planned Versus Actual Time to Complete Work	2
Budget Variance	3
Percent of Staff with Met Target Proficiency Level	2

Table 6 Example of Measure Purpose Values

Measure Development Evaluation

The second set of criteria in our process will help you assign a value for the feasibility level of developing each measure candidate. This process is valuable because it forces you to think about, and possibly research, the source of your measure data before you are committed to producing that measure. This evaluation also provides documentation of the effort required for a specific measure so that you may effectively communicate measure availability to your stakeholders. Below you will find descriptions and examples of the Measure Development Criteria: Organizational Readiness, Data Availability, Frequency and Consistency, Respondent and Organizational Burden, and Effort to Analyze and Format Data.

Organizational Readiness

We define organizational readiness in terms of the maturity of the organization's processes in place to capture measure data. Although it is possible to gauge the performance of an organization operating with poorly formed or undocumented processes, such a measurement is likely to yield only gross information. Obtaining the detailed information necessary to focus the organization on performance improvement requires data collection processes and well understood definitions for the data responses. See Table 5 below for examples of how to assess organizational readiness.

Measures	Organizational Readiness
Percent of Research Proposals Accepted by Customer	Processes are in place to control the development of research proposals and track sponsor acceptance
Percentage of Change Requests Implemented	Processes are well defined, staff turnover has led to loss of trained PMs
Planned Versus Actual Time to Complete Work	Processes are in place to capture resource time, staff are trained and required to report their time
Budget Variance	Processes are in place to capture budget expenditures and for setting budget plan
Percent of Staff with Met Target Proficiency Level	Process and instrument are in place to collect data, staff are trained on platform

Table 5 Example of Organizational Readiness Responses

Data Availability, Frequency and Consistency

We define data availability as both the existence of the data and the ability to access the data. This criterion may seem like operational readiness, but it is there to capture situations where you may not have access to data, or the data has yet to be collected. We included frequency to capture a situation where the data is not collected frequently enough to be useful for your measure. We included consistency to capture any situations where the data collected are not well defined, well understood or there has been a change in definitions in the data series. See Table 6 below for examples of how to assess data availability, frequency, and consistency.

Measures	Data Availability, Frequency and Consistency
Percent of Research Proposals Accepted by Customer	Available, annual, well defined
Percentage of Change Requests Implemented	Available, On Demand, Well Defined
Planned Versus Actual Time to Complete Work	Data is stored on database, data is available by day, collection platform produces consistent data
Budget Variance	Data is mostly reliable; data may not be available at project level
Percent of Staff with Met Target Proficiency Level	Data is downloadable, recent data is not available, platform produces consistent data

Table 6 Examples of Data Availability, Frequency, and Consistency Responses

Respondent and Organizational Burden

We define respondent burden as the effort it takes for staff to provide measure data and more specifically, how staff perceive the burden of providing measure data. This can be as simple as filling out a timesheet or completing a questionnaire. We define organizational burden as the effort it will take for an organization to generate the measure data. If a process is in place to collect or generate the data and staff are available and trained, then the burden/difficulty would be low. Medium effort could be defined as a more time-consuming process such as compiling data from multiple systems and spreadsheets or requesting data from other staff. If a system must be designed and developed and staff must be trained, then the burden/difficulty would be much higher. See Table 7 below for examples of how to assess respondent and organizational burden.

Measures	Respondent and Organizational Burden
Percent of Research Proposals Accepted by Customer	Staff are socialized to develop research ideas into formal proposals, organization has provided resource time to develop and manage research proposals
Percentage of Change Requests Implemented	New staff need to be trained on process; organization will need to find resources for training
Planned Versus Actual Time to Complete Work	Staff is socialized to provide input; organization has invested resources into developing and implementing collection process
Budget Variance	Organization has invested resources to put processes in place to collect cost data, need to estimate budget data if costs not captured at project level
Percent of Staff with Met Target Proficiency Level	Organization has invested resources to put processes in place, Staff will have to go in and update their responses

Table 7 Examples of Respondent and Organizational Burden Responses

Effort to Analyze and Format Data

We define the effort to analyze and format data as the level of difficultly required to format the data from its current form to meaningful measure data and the level of difficulty required to interpret the data into an accurate representation of the situation. This criterion can and should be influenced by the technical skill available to produce the measures. If you have a person that is experienced with accessing and formatting data, then the difficulty would be low. If you have someone with only a basic knowledge, then the difficulty would be high. Additionally, if you have too much data, too little data, or data values with a very large spread, then this too would influence how you rate this criterion. See Table 8 below for examples of how to assess effort to analyze and format data.

Measures	Effort to Analyze and Format
Percent of Research Proposals Accepted by Customer	Research proposals are on SharePoint site which produces measure data in Excel format
Percentage of Change Requests Implemented	Some effort required to format data into final metric
Planned Versus Actual Time to Complete Work	Some effort and expertise required to retrieve data from database and format data into measure
Budget Variance	Data will need to be keyed into new measure, may need to estimate project budget data if costs are not captured at the project level
Percent of Staff with Met Target Proficiency Level	Data will need to be formatted into new measure

Table 8 Examples of Effort to Analyze and Format Data Responses

Scoring of Measure Development Criteria

The second step of the measure development evaluation process is to assign values to the details you captured about each criterion. For determining the values, we focused on the "feasibility" of development and Table 11 below shows the values in the scale. We have found that a scale of

three values: easy, moderate, and hard, is enough to produce a spread of values that is meaningful in the final scoring process. We have found that having more "middle values" only adds confusion to the scoring process. You can add weights to the criteria values, but they need to be large enough to make a significant difference. Table 12 below shows the values we have assigned to each of the criteria. Note that we have added up the values to produce a total for the measure development score.

Value	Feasibility Level
1	Hard
2	Moderate
3	Easy

Table 11 Measure Development Ranking Value Definitions

Measures	Organizational Readiness	Data Availability, Frequency and Consistency	Respondent and Organizational Burden	Effort to Analyze and Format	Total Score
Percent of Research Proposals Accepted by Customer	3	3	3	3	**12**
Percentage of Change Requests Implemented	2	3	2	3	**10**
Planned Versus Actual Time to Complete Work	3	3	3	2	**11**
Budget Variance	3	2	2	2	**9**
Percent of Staff with Met Target Proficiency Level	3	2	2	2	9

Table 12 Example of Feasibility Evaluation Values

Final Scoring of the Measures

Now that all the hard work is done, we generate a final score by combining the development score and the purpose score. In table 13, we have the formula for the final score which is a straight multiplication of the development score by the purpose score. Table 14 shows the results of combining the two scores.

We purposely call this a scoring process and not a ranking process because it is the combination of scores that will provide the final values that can then be used to make decisions. The scores highlight some different results based on the combinations. On table 14, the front runner on the development score, Percent of Research Proposals Accepted by Customer, stays on top with a high purpose score. Whereas a measure with a low development score, Budget Variance, is elevated to second place by a high purpose score. These score combinations help to show the

"low hanging fruit" measures that can be quickly put in place and the heavier lift of some important but more difficult to develop measures.

Development Score	X	Purpose Score	=	Final Score

Table 13 Formula for Final Score

Measures	Development Score	Purpose Score	Final Score
Percent of Research Proposals Accepted by Customer	12	3	**36**
Percentage of Change Requests Implemented	10	1	**10**
Planned Versus Actual Time to Complete Work	11	2	**22**
Budget Variance	9	3	**27**
Percent of Staff with Met Target Proficiency Level	9	2	**18**

Table 14 Matrix of Assigned Evaluation Ranking Values and Final Scores

Final Considerations for Your Measure Program

We hope you have found our reality check process enlightening and a promising application for your work environment. Implementing a performance management program can be as challenging to implement as it is beneficial to your program management and with that knowledge, we would like to leave you with some questions to think about when developing and managing a measure program.

Priorities - How much support do you have in your organization to keep producing the data needed by the measures? Will that continue with a change in leadership?
Initiation – Who has the authority to initiate your measure program? When do they need for the measures to be up and running?
Resources - What resources do you have to develop the measures? What resources will you have to maintain the measures and dashboards? What resources do you have to report the measures?
Platform – Where will you house your measures? What resources do you have to develop and maintain a platform for developing and reporting your measures?
Maintenance - Will your organization be committed to reviewing the data and updating the measures? What access will your stakeholders have to the measures? How frequently do you need to provide measure data?
Updates - Does the measure point of view still resonate with stakeholders? Should the measure be expanded or contracted? Are any of your measures outdated and ready for retirement?