Performance management, managing for results, or managing for outcomes . . . is now one of the most widely adopted management practices,” Richard M. Walker, Chan Su Jung, and George A. Boyne note in their article, “Marching to Different Drummers? The Performance Effects of Alignment between Political and Managerial Perceptions of Performance Management.”

This should hardly be surprising. Performance management, the use of goals and measurement as management tools, makes intuitive sense. Who can argue with the logic of setting goals, especially goals that are clearly linked to improving desired outcomes? Goals, succinctly communicated by managers, signal what, how much, and how soon to the people in an organization (including the boss, subordinates, and other offices), delivery partners (e.g., contractors, grantees, and other government agencies), authorizers (e.g., legislators), and others. Goals aligned with priorities and available resources effectively focus the use of organizational resources, including energy and intelligence.

Similarly, who can argue with a manager’s need for multiple types of measurement to gauge the impact of agency action, especially when complemented by analysis, data-driven follow-up questions, and occasional evaluations to isolate the impact of agency action from other variables influencing outcomes?

Of course, government agencies do not always manage with clear goals and robust measurements, but few would argue that, in most cases, they should. Therefore, it is great, as Walker, Jung, and Boyne observe, that performance management has become a “critical area of enquiry.” Practitioners can benefit significantly from research that reveals how significant changes in outcomes result from variations in performance management practice (e.g., how data analyses are presented and to whom they are disseminated; with whom and how performance information is discussed; measurement frequency; how goals and measurements are linked to penalties and rewards).

Before such analysis can have value to practitioners, however, researchers need to find valid means to measure performance, as Stéphane Lavertu, David E. Lewis, and Donald P. Moynihan’s research reminds us in “Government Reform, Political Ideology, and Administrative Burden: The Case of Performance Management in the Bush Administration.” Lavertu, Lewis, and Moynihan conclude that the Program Assessment Rating Tool (PART) scores that the George W. Bush administration assigned to programs
as an indicator of performance were subjective. PART scores were the opinion of a single budget examiner, inevitably risking a ratings bias that skating and gymnastics scoring long ago tried to address by using multiple raters and sometimes eliminating high and low scores. The authors find that PART scores also reflected the biases of the examiners’ political leaders, answering a question raised by Walker, Jung, and Boyne: government career staff, at least in the U.S. federal budget office during the period studied, aligned their ratings with the preferences of their political bosses.

In truth, PART scores were never an indicator of agency performance. Rather, they assessed agency performance management practice, summing a checklist of 25 yes/no questions. The best budget examiners used PART scores to strengthen agency measurement and evaluation practice, but the scores never worked well as a performance indicator (Metzenbaum 2009).

Lavertu, Lewis, and Moynihan’s findings should warn the many researchers who are tempted to grab PART scores or other subjective measures as proxies for performance: don’t do it. Weak performance measures used as dependent variables compromise the validity of performance management research findings.

Walker, Jung, and Boyne use what, at first, appears to be a more promising dependent variable, a composite score of six aspects of organizational activity. But while parts of the composite score could be valid as a performance indicator (e.g., average school passes at 16), the score can no longer convey performance information when combined with activity counts. The problem is exacerbated by a weighting scheme that uses a subjective determination of the relative import of different budget areas. Managers may be able to use the composite score fruitfully to trigger focused follow-up questions (e.g., are high scorers doing something worthy of replication, do low scorers need attention?), but the composite score does not work well as a performance indicator.

These two papers make abundantly clear that performance management researchers need valid performance measures for valid conclusions. The good news is that government agencies are getting better at measuring performance, so more data should be available for interested researchers.

That is not to suggest that the data are always easy to access for analysis. Government agencies and researchers need to find smarter ways to collect and share data with those willing to do deeper analyses, both in academia and in government. Government and researchers willing to run objective iterative, measured comparisons of different program practices would also benefit from closer relationships. In addition, researchers could help government find smarter ways to assess and compare the impact of different agency actions in hard-to-measure programs such as drug-running prevention, where what needs to be measured tries to hide, and basic science with highly uncertain but potentially high-value returns.

One final note: Walker, Jung, and Boyne assert that performance management is based on principal–agent theory, involving goals, targets, indicators, and “taking action on the basis of whether the targets are achieved.” The presumption that an “action” (presumably a reward or punishment) will be taken depending on whether or not targets are achieved has proven to be a potentially dangerous paradigm, as Kettl hinted in 1997. A direct linkage between target attainment and incentives often plays out in practice with a retreat to timid targets, compromising the innovation-inspiring value of a stretch goal, or measurement manipulation, necessitating more spending on monitoring. A direct linkage also tends to undermine other powerful employee motivators, such as intrinsic drive (see Metzenbaum 2006; Pink 2009 for research citations). Further, it can divert scarce analytic resources from important problem-solving tasks such as searching for root causes and effective treatment to less valuable endeavors such as negotiating regression-adjusted targets ( Heckman et al. 2011). But perhaps the biggest problem is that it mistakenly suggests that the true objective of performance management is hitting a target rather than improving performance and increasing public-value return on investment.

Many of us working in and with government are trying hard to reset this mistaken mind-set, treating target attainment as the purpose rather than a means to an end. It is my hope that researchers, in choosing areas and methods of study, will redirect their inquiries to the real purpose of performance management: continually finding and applying government practices that work better (Gawande 2007).

References


