A FRAMEWORK FOR EVALUATING THE EFFECTIVENESS OF PERFORMANCE MEASUREMENT SYSTEMS

Lori Criss Powers
March 28, 2009

Abstract: Performance measurement systems are widely used by organizations in both the public and private sectors, with the objective of improving decision making and performance. The cost associated with building and implementing these systems is high, yet many fail to deliver the expected results. This paper proposes an evaluation method and framework using an evidence-informed theory of change. The proposed theory of change can be used to guide the evaluation of the effectiveness of implemented performance measurement systems or to inform the design of new systems.

Key Words: Performance measurement systems, evaluations of effectiveness, theories of change, logic models.

Recommended Citation: Powers, L.C. (2009). A framework for evaluating the effectiveness of performance measurement systems. RealWorld Systems Research Series 2009:1. Available at SSRN: http://ssrn.com/abstract=1371158

INTRODUCTION

There are very few systematic evaluations of the effectiveness of performance measurement systems reported in the literature, and while convincing proof of the value of these systems remains elusive, more and more problems and system failures are being reported (e.g., Bourne, Kennerley & Franco-Santos, 2005; Perrin, 1998; Perrin, 2003; Santos, Belton & Howick, 2002; USGAO, 2000; Wholey, 2002). Despite this state of affairs performance measurement systems continue to be widely implemented by organizations in the public, non-profit and private sectors throughout the world.

With the overall objective of improving decision making and performance (Johnston, Brignall, & Fitzgerald, 2002; Plant & Douglas, 2006; Poister, 2003; USGAO, 2001; Wholey, 2002) these systems, though difficult and expensive to implement (Johnston, Brignall, & Fitzgerald, 2002; Plant & Douglas, 2006) are increasingly seen as an important and indispensable management tool (Plant, Agocs, Brunet-Jailly & Douglas, 2005). And, though several drivers influence management decision making, especially in complex environments (Perrin 1999; Poister, 2003; Sheirer & Newcomer, 2000; Liner et al., 2001), performance measurement data is often expected to act as a key source for informing the improvement efforts of numerous management functions such as strategic planning and management, operational planning and management, communications with stakeholders, budgeting, human resource management, evaluation, and so on (Poister, 2003).

These systems continue to be implemented based on the assumption that they will have a positive impact on performance (Bourne, Kennerley & Franco-Santos, 2005). Even organizations with immature and ad hoc management processes that lack the capacity to build or use them properly (Plant, Agocs, Brunet-Jailly & Douglas, 2005; USGAO, 2001), including many in the not-for-profit sector, will often have some kind of performance measurement system in place because of the pressures imposed by funders and governments to demonstrate acceptable performance (United Way of America, 2000; Wholey, 2002). For example, the Government Performance and Results Act forced agencies in the US federal government to adopt a results focus and thus performance measurement systems. To satisfy their own need for outcome data these agencies in turn began pressuring non-profit service providers and others to also adopt a results-based focus and thus performance measurement systems (Hatry & Lampert, 2001).

At the same time, there is continuing confusion and debate over how best to design these systems (e.g., Bourne, Kennerley & Franco-Santos, 2005; Kueng, Meier & Wettstein, 2001), how best to measure performance (e.g., Perrin, 1998; United Way of America, 2000) and how best to report (e.g., Auditor General of Canada, 2002; USGAO, 2000; Wilkins, 2004) and use the results (e.g., Johnston, Brignall, & Fitzgerald, 2002; Plant, Agocs, Brunet-Jailly & Douglas, 2005).

Given these realities, organizations and evaluators are beginning to ask some key questions: Do these systems provide value (Johnston, Brignall, & Fitzgerald, 2002)? Are the limitations of performance measurement systems mainly related to poor system design (Kueng, Meier &

Wettstein, 2001) or inaccurate or misleading data and faulty reporting (USGAO, 2000)? Or, is there a basic flaw in the logic of how performance measurement systems work (Perrin, 1998)?

To answer these questions this paper advocates that performance measurement systems be systematically evaluated for their effectiveness. But one question remains, how can we judge the effectiveness of a performance measurement system with the lofty and difficult to operationalize goals of improved decision making and improved performance?

This paper proposes an evaluation method and framework, using an evidence-informed theory of change illustrated in a logic model, to help answer this question.

A logic model is an attempt to capture the conceptual relationship between long term outcomes and the interventions being implemented so that organizations can demonstrate the probable effectiveness of their activities (Hatry & Newcomer, 2004; Sheirer & Newcomer, 2000; Weiss, 1998). If there are strong logical and/or proven links between the activities and the desired outcomes, there is no need to measure the long term outcomes separately (Weiss, 1998).

This paper suggests that evaluators using an evidence-informed theory of change to frame the evaluation of a performance measurement system should be able to assess system effectiveness with a fair degree of confidence. As well, use of the framework and methodology provided here should enable an evaluator to make evidence-informed recommendations for the improvement of a performance measurement system during any stage of its life-cycle. Furthermore, since performance measurement system data often are intended to inform program, organizational or policy level evaluations, the evaluation and subsequent improvement of the measurement system itself should strengthen the connection between the two. Strengthening the relationship between evaluation and performance measurement is an important goal that is supported by many in the evaluation community (e.g., Sheirer & Newcomer, 2000; Wholey, 1999)

The remainder of this paper is organized into four sections as follows: [1] *Methodology*: This section discusses the methodology used to develop this paper [2] *Developing the theory of change for a performance measurement system*: This section discusses terminology, proposes a boundary for the performance measurement system and presents, in a step-by-step fashion, how and why the component parts of the theory were developed [3] *Putting it all together:* This section illustrates the full theory of change by putting the component parts discussed in the previous section together [4] *Planning and implementing the evaluation using the theory of change*: This section offers detail around how the theory of change might be used to plan and implement an effectiveness evaluation of a performance measurement system. The section also briefly discusses how issues peripheral to a performance measurement system might impact its effectiveness and how the evaluator might deal with these issues [4] *Conclusions and implications:* The paper closes by summarizing its conclusions, implications and limitations.

METHODOLOGY

This paper is based on a literature review of over one hundred and fifty peer-reviewed articles and books that concerned performance measurement systems. The literature included items related to performance measurement systems for non-profit, government and for-profit organizations. Perspectives from numerous disciplines were represented, including those from management, accounting and finance, public administration, systems, organizational development and evaluation.

The literature was reviewed in two broad phases. The first phase investigated the different approaches that have been used to judge the effectiveness of performance measurement systems and to decide if a theory of change method might be appropriate. This phase of the review found that of the publications located, only a small portion dealt specifically with ways to assess the effectiveness of performance measurement systems and none used a theory of change in their approach - though one logic model for performance measurements systems adapted from McLaughlin and Jordan (2000) was kindly supplied by the authors (see McLaughlin, J. A., & Jordan, G. B. in Wholey, Hatry, & Newcomer, 2004). That logic model, while helpful from a descriptive standpoint, was insufficiently developed for the purpose of evaluating effectiveness.

This paper does not offer a critique of the approaches used to assess the effectiveness of performance measurement systems as identified in the literature, instead this paper proposes a new approach. Readers interested in assessing these other approaches may wish to begin with a review of the following sources: Bourne, Kennerley & Franco-Santos, 2005; Holloway, 2001; Kennerley & Bourne, 2003; Najmi, 2005; Poole, Nelson, Carnahan, Chepenik & Tubiak, 2000; Rivenbark & Roenigk, 2006; Tangen, 2004; Van Aken et al., 2005.

The second and larger phase of the literature review focused on developing a systematic approach for the creation of a theory of change for performance measurement systems as well as locating evidence to inform the development of such a theory. Findings from this portion of the review are given throughout this paper. The number of citations is limited to those judged necessary to support a given point and to those that are more or less typical of the overall literature. Readers interested in obtaining the full list of papers reviewed for this study are invited to contact the author for more information.

Lastly, it is noted that there has been little quantitative research conducted on performance measurement systems. The data presented in the literature tends to be qualitative and often derived from case studies. Nevertheless, qualitative data is highly compatible with the theory of change concept (Weiss, 1998) and is used here without compunction.

DEVELOPING A THEORY OF CHANGE FOR A PERFORMANCE MEASUREMENT SYSTEM

Identifying the System Boundary

The terms performance measurement (or monitoring) system and performance management (or results-based management or performance-based management) system are often used interchangeably in the literature. For example, Wholey (2002) with an eye on the public sector uses the term 'performance-based management' and defines it as "the purposeful use of resources and information to achieve and demonstrate measurable progress toward agency and program goals". Neely, Gregory & Platts (1995) with an eye on the private sector use the term 'performance measurement system' and define it as "the set of metrics used to quantify both the efficiency and effectiveness of actions". The Treasury Board Secretariat of Canada (2006) uses the term 'results-based management' (RBM) and defines it as "a life-cycle approach to management that integrates strategy, people, resources, processes and measurements to improve decision-making, transparency, and accountability. Poister (2003) uses the term 'performance measurement system' and defines it as a "management system that tracks selected performance measures at regular time intervals so as to assess performance and enhance programmatic or organizational decision making, performance, and accountability". Plant, Agocs, Brunet-Jailly & Douglas (2005) influenced by systems theory, use the term 'performance management system', and define it as "an organizational concept that is comprised of systems of organizational goals, processes and relationships".

The problem with these definitions is not only the confusion caused while reading the literature - the real problem is that these definitions suggest different boundaries for the system (Williams & Imam, 2006; Yourdon, 1989). For example, results-based management (or performance management) systems focus on indicators (or measures) as the primary source of input into the management process with occasional evaluations to provide 'complementary' information (Development Co-Operation Agencies, 2000; Treasury Board Secretariat of Canada, 2006). Sheirer & Newcomer (2000), Perrin (1999), Poister (2003), and Wholey (2002) on the other hand, seem to expand the boundary to include other sources of performance information (e.g., findings from audits, case studies, etc.).

From a systems perspective, it is necessary to understand the system's boundaries in order to determine what the associated activities, inputs, outputs and outcomes to the system should theoretically be. This paper adopts the view of performance measurement systems as given by Wholey (2002). Wholey's view is that "performance information may come from performance measurement systems, audits, case studies, benchmarking comparisons, basic and applied research, program evaluations, and experiments".

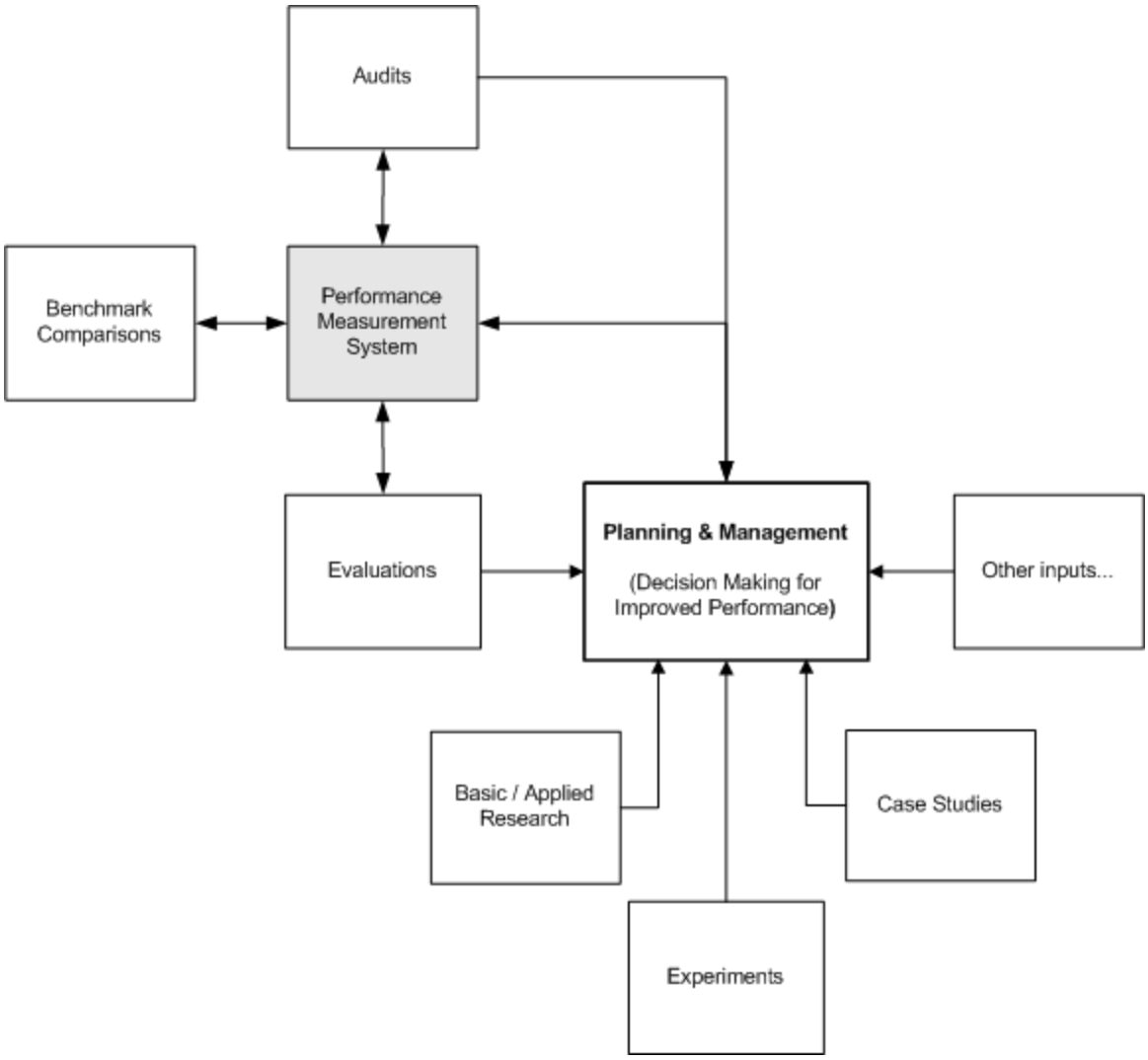

Fig 1: Wholey's (2002) conceptualization (adapted) of where the performance measurement system fits within the context of organizational planning & management

Figure 1 illustrates Wholey's view and adapts it in two ways. First, double-sided arrows are used to show the relationship between the inputs and the 'back and forth' nature between measurement and management. Second, a placeholder was added for the 'other inputs' that may impact on how and why decisions are taken within an organization. This study did not focus its investigation on these 'other inputs' but it is hypothesized that they might include the past experience of the decision maker(s), political pressures exerted on the decision maker, and so on. Whatever these other inputs may be, this view helps us to establish the system boundaries by demonstrating that performance measurement systems provide only one source of information to the decision maker. An alternate view might place the performance measurement system at the 'center' where all inputs are accounted for and captured within the system. This paper adopts the first view (as illustrated in Figure 1) primarily because the reality in most organizations is that performance measurement systems are just not that well built or

mature enough to support the view that they are, or should be, at the 'center' of the management framework.

Identifying the Activities Associated with a Performance Measurement System

In order to evaluate the effectiveness of a performance measurement system using a theory of change it is necessary to identify the activities associated with a system of this type. Several perspectives from the literature were helpful for this purpose.

The activities that always occur (within any performance measurement system) were identified first. For this, the systems view was most helpful. For example, Kueng, Meier & Wettstein (2001) as well as Bourne, Mills, Wilcox, Neely & Platts (2000) reflect on performance measurement systems in terms of their life-cycles and broadly describe the following activities:

- Design (the system is planned and described)
- Build and Implement (the system is constructed and tested and procedures, to collect data for example, are put into place and the system is deployed)
- Run or Use (the system is operational)

Kueng, Meier & Wettstein (2001) also give us a decommissioning activity (when the system is retired or uninstalled); however, this activity is outside the scope of interest for this paper and so was not included in the logic model.

It is important to note at this point that the activities listed above are conceptual, that is, from a practical standpoint they overlap and there is not a simple linear sequence from one activity to another (Bourne, Mills, Wilcox, Neely & Platts, 2000). This concept is captured in the logic model (as shown in Figure 5) and is illustrated through the use of double-sided arrows between the activities that tend to have multiple iterations (between design and build for example).

Additional core activities identified in the literature expand on the 'run' phase. These include:

- Data collection (e.g., Wholey, Hatry, & Newcomer, 2004)
- Data quality control (e.g., Perrin, 2003; USGAO, 2000)
- Performance data analysis and reporting (e.g., Auditor General of Canada, 2002; USGAO, 2000)
- Feedback (for maintaining and improving the system) (e.g., Ernst, K. 2002; Franco-Santos & Bourne, 2005; Henri, 2004; Smith & Goddard, 2002; Liner et al., 2001)

In an ideal performance measurement system the activities given above (i.e., design, build, test, deploy, data collection, data quality control, data analysis and performance reporting and feedback) will always occur.

Next activities that *may or may not* be associated with a particular performance measurement system were identified. The managerial activities view was most helpful for this purpose.

Poister, (2003), for example, describes ten managerial functions that performance measurement systems are intended to support in public and non-profit organizations. These are:

- Monitoring and reporting
- Strategic planning
- Budgeting and financial management
- Program management
- Program evaluation
- Performance management (in terms of directing and controlling human resources)
- Quality improvement, process improvement
- Contract management
- External benchmarking
- Communication with the public

Poister's list is intended to be comprehensive, though some might argue that it is incomplete. However, it serves to demonstrate the number and variety of activities that performance measurement systems could incorporate. In any case, the logic model developed here purposely does not include every possible 'specific use' - only *some* of the more common specific use activities are included. The evaluator of a performance measurement system using this framework in the field to guide an effectiveness evaluation will need to flesh out the specific use activities for the particular organization and system under study (this topic is discussed further later in this paper). The logic model developed here includes an 'other' activity intended to serve as a placeholder for these 'to be determined' activities.

Identifying Desired Medium and Long Term Outcomes

Obviously, the evaluator of a performance measurement system needs to define the system's intentions before beginning an assessment; otherwise he/she will not be able to judge system effectiveness. In practice, surfacing the mid- to long-term outcomes for a performance measurement system does not appear all that difficult. Performance measurement is fundamentally for the purpose of improving organizational outcomes through improved decision making (Plant, Agocs, Brunet-Jailly & Douglas, 2005, Poister, 2003). For example, the following sample list of mid to long term outcomes extrapolated from the literature can be adapted to fit the situation under study:

- Performance information is used to achieve performance goals (Wholey, 2002)
- The organization focuses on effectiveness and efficiency and continuously improves its programs and operations and seeks approaches to maximize limited resources (USGAO, 2001)
- Desired organizational results are achieved within acceptable, predefined ranges of economy, efficiency and effectiveness (Reider, 2001)

The real challenge for the evaluator comes when trying to understand the intended short term outcomes of a particular performance measurement system. The problem is that often the system's owners or stakeholders do not readily know or cannot readily articulate precisely what the system is supposed to accomplish in the short term (Kueng, Meier & Wettstein, 2001). Should the system be used to inform strategic planning? Should it be used to manage staff performance? Should it be used to improve the quality of services or products (Kueng, Meier & Wettstein, 2001)? Ideally, these are questions that should be asked and answered while the system is being designed. Indeed, it can be argued that not knowing *precisely what* the system outcomes should be is *precisely why* these systems tend to fail to live up to expectations.

Let's take a moment to review what has thus far been accomplished regarding the development of the theory of change. We have identified the activities that must always occur for a performance measurement system; the activities that may or may not occur; midterm outcomes; and long term outcomes. Next we identify inputs, outputs and short term outcomes for each of the activities.

Identifying Inputs, Outputs and Short Term Outcomes

To identify the sample inputs, outputs and short term outcomes the findings from the literature were reorganized and problems and promising practices associated with each performance measurement system component (i.e., design and implementation, run, performance information utilization) were identified and captured in tabular format. Because issues were not necessarily reported according to these components (categories), certain issues were placed into more than one component (category). Not every issue identified in the review, nor every author reporting an issue was captured in the table. Only those that appeared generalizable across sectors and were more or less typical of the overall literature were included.

Through careful analysis the issues in the table were then grouped into themes. The themes were used to extrapolate a set of sample inputs, outputs and outcomes for the logic model. Table 1 shows the table used for this purpose:

COMPONENT / THEME	EXAMPLE SOURCES
Design and Implementation	
Performance measurement systems are poorly designed, built and tested	Kueng et al., 2001; Plant et al., 2005; Poister, 2003
Performance measurement systems are poorly managed	Bititci et al., 2000; Franco-Santos & Bourne, 2005
Performance measurement systems do not meet the needs of users	Kueng et al., 2001; Perrin, 2003; Plant et al., 2005; Poister, 2003
Organizations do not have the capacity to collect and monitor performance data	Kueng et al., 2001; Plant et al., 2005; Poister, 2003
Run	
Data collection procedures are ad-hoc	Liner, et al., 2001; Poister, 2003; USGAO, 1999
Data collected is not relevant or useful	Plant et al., 2005
Data collection costs are too high	Johnston et al,, 2002; Kueng et al., 2001; United Way of America, 2000
Data validation and verification procedures are inadequate / regular audits are recommended	Bernstein, 2001; Perrin, 2003; USGAO, 2000
Data collected is of poor quality	Perrin, 2003; Poister, 2003; USGAO, 2000; Wholey, 1999
Performance reports are not usable	Auditor General of Canada, 2002; Kueng et al., 2001; Liner, et al., 2001; Perrin, 2003; Plant et al., 2005; USGAO, 2000, 2001; Wilkins, 2004; Wholey, 1999
Performance Information Utilization	
Performance data is not used effectively	Franco-Santos & Bourne, 2005; Liner, et al., 2001; Plant et al., 2005; Poister, 2003; USGAO, 2000
System feedback mechanisms are inadequate	Bititci et al., 2000; Franco-Santos & Bourne, 2005; Henri, 2004; Smith & Goddard, 2002; Liner et al., 2001;
External Factors	
External factors effect performance	Auditor General of Canada, 2002; Plant et al., 2005; Poister, 2003

Table 1: The table used to extrapolate the set of inputs, outputs and short term outcomes

The following discussion provides three examples, one for each of the major performance measurement system components (i.e., design and implementation, run, performance information utilization) demonstrating how the analysis of the issues identified in the literature and captured in the table shown in Table 1 were used to extrapolate a set of sample inputs, outputs and outcomes for the logic model.

Issues Related to the Design and Implementation Component

Example 1: Analysis of the issues and themes led to an arguably obvious but important conclusion - most performance measurement systems are simply not designed properly. Improperly designed performance measurement systems will often have unclear goals and target audiences, leading to questions like: Which level of aggregation is useful? To whom should the performance results be communicated? What should recipients do with the information they get (Kueng, Meier & Wettstein, 2001)?

Looking at the full set of findings, it was hypothesized that a properly engineered performance measurement system would avoid many of the most common problems associated with the system design activity and thus would avoid many other problems downstream in the life-cycle. Thus a desired short term outcome was identified: namely, a performance measurement system should be developed in accordance with sound system engineering principles.

Looking then at the systems literature it was determined that, among other things, successful systems development efforts will document stakeholder needs and required system functionality early in the design phase (Yourdon, 1989). Thus the outputs for the design activity are now identified to include, for example, a 'systems requirements document' that defines stakeholder needs. A corresponding input is then easily identified as; 'stakeholder consultations'. Note how these inputs and outputs relate to the issues and themes identified in the literature. For example for the 'performance measurement systems do not meet the needs of the users' issue (e.g., Kueng, Meier & Wettstein, 2001) - the proper development of a systems requirement document addresses this problem because the requirements document forces consultations with the stakeholders to identify their needs.

Figure 2 illustrates the sample inputs, outputs and short term outcomes that resulted from the complete analysis of the design and implementation component and its related issues:

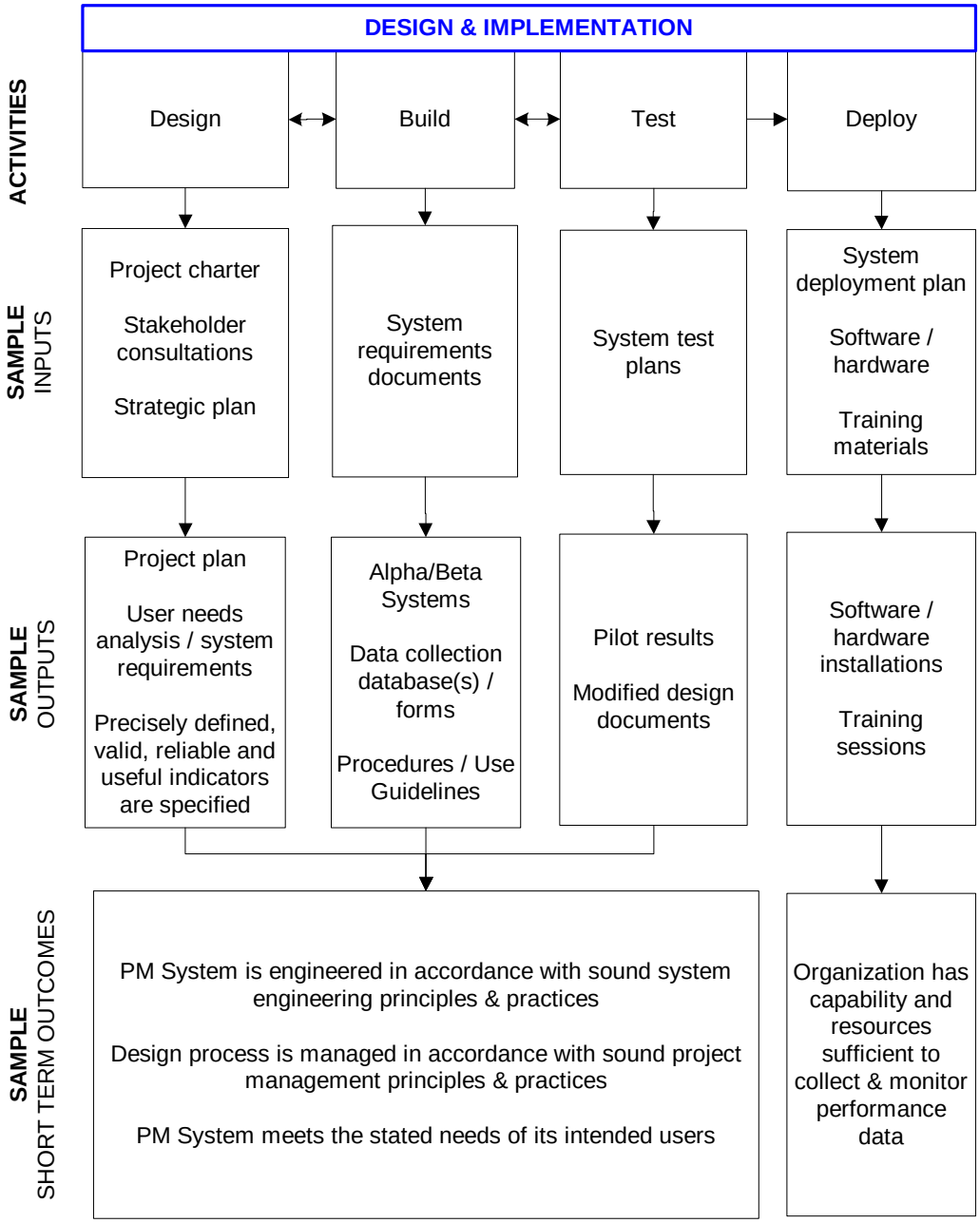

Fig 2: The Design & Implementation Component of the Theory of Change

Issues Related to the Run Component

Example 2: As can be seen from the findings listed in Table 1 the literature identifies a common and critically important problem with many performance measurement systems - the data collected and used to develop performance reports are often not seen as credible by the users of those reports. This lack of credibility leads to a number of lethal issues for a performance measurement system, chief among them that the data will likely not be used in an effective manner, if at all (e.g., Poister, 2003). A promising practice recommendation is that regular

audits and other data quality control techniques be used to validate and verify the data (e.g., Bernstein, 2001; Perrin, 2003; USGAO, 2000) thus improving its quality and credibility. The logic model proposed in this paper therefore includes inputs, outputs and outcomes related to this issue (e.g., input to the data quality control process includes audits) as illustrated in Figure 3 below.

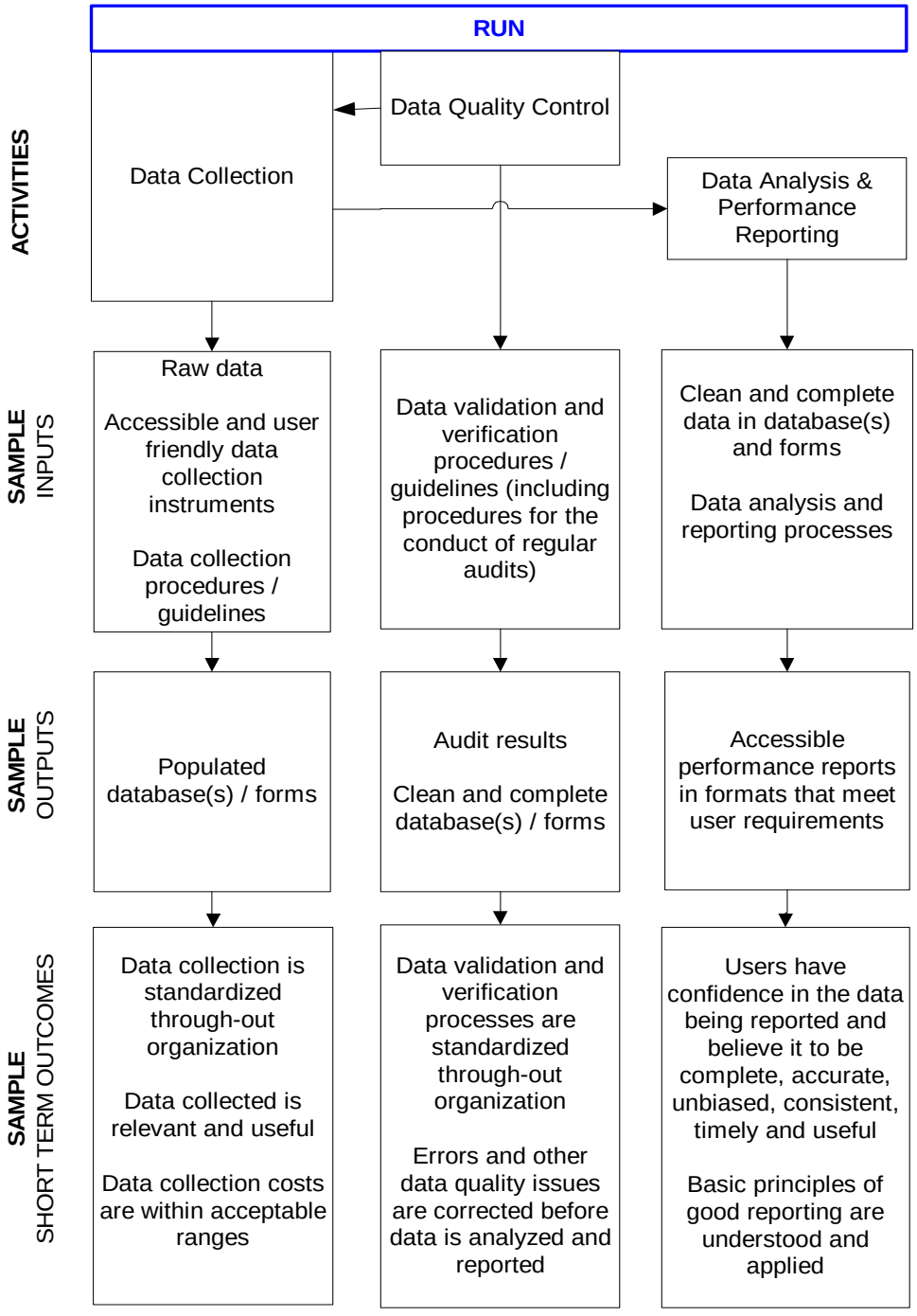

Fig 3: The Run Component of the Theory of Change

Issues Related to the Performance Information Utilization Component

Example 3: Identifying issues related to the performance information utilization component from the literature was more difficult than for the other components. Though it was possible to generalize some findings, many issues reported in this category are highly contextual. This study found that searching the literature for promising practices associated with specific uses such as strategic planning, budgeting and so on was at least helpful for the development of the logic model inputs, outputs and outcomes. For example, the United States General Accounting Office (e.g., USGAO, 2000; 2001) has had arguably more experience than any organization in the world in dealing with performance reporting and the budgeting function. The specific problems they report regarding specific agencies and their budgeting functions may or may not be germane to all organizations, but when their careful investigations lead them to state that including a description of performance data limitations in planning/performance reports increases data credibility and therefore leads to increased use of the performance information then the practice should be seen as at least potentially promising. This promising practice, supported by other information gathered from the literature, led to the inclusion of this outcome for the budgeting & financial management activity that is the "organization produces reliable budget estimates and relates performance budget, spending and workforce information in a credible and useful manner".

Figure 4 illustrates the set of sample inputs, outputs and outcomes for the performance information utilization component of a performance measurement system (note that the evaluator will need to expand the list of inputs, outputs and outcomes for the specific use activities - this issue is discussed in a subsequent section):

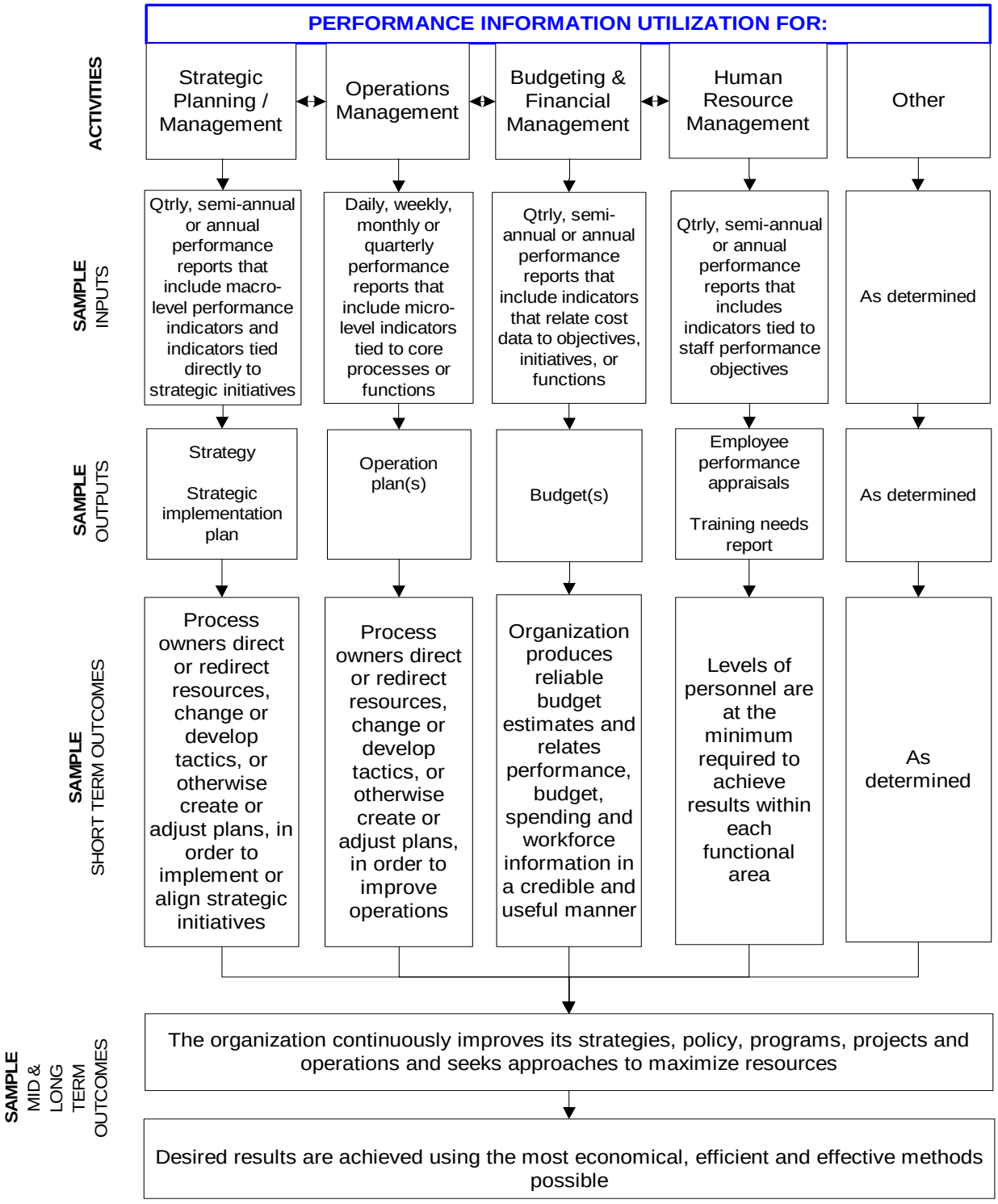

Fig 4: The Performance Information Utilization Component of the Theory of Change

PUTTING IT ALL TOGETHER: A THEORY OF CHANGE FOR PERFORMANCE MEASUREMENT SYSTEMS

The following logic model diagram places the component parts of the theory discussed above (Design & Implementation, Run and Performance Information Utilization) together into a single diagram in order to illustrate the entire model (to keep it legible the diagram omits the inputs and outputs that were shown above in Figures 2, 3, and 4).

It is important to note at this point that the theory of change as presented here does not, as Weiss (1998) emphatically states: "need to be uniformly accepted. It doesn't have to be right" in order for it to be useful. In fact, this theory is just one explanation of how a performance measurement system might work effectively. The hypothesis is that if the "activities are conducted as planned; with sufficient quality, intensity, and fidelity to plan the desired results will be forthcoming" (Weiss, 1998).

The illustration is followed by a detailed discussion that describes how the model could be used for evaluating performance measurement systems in practice.

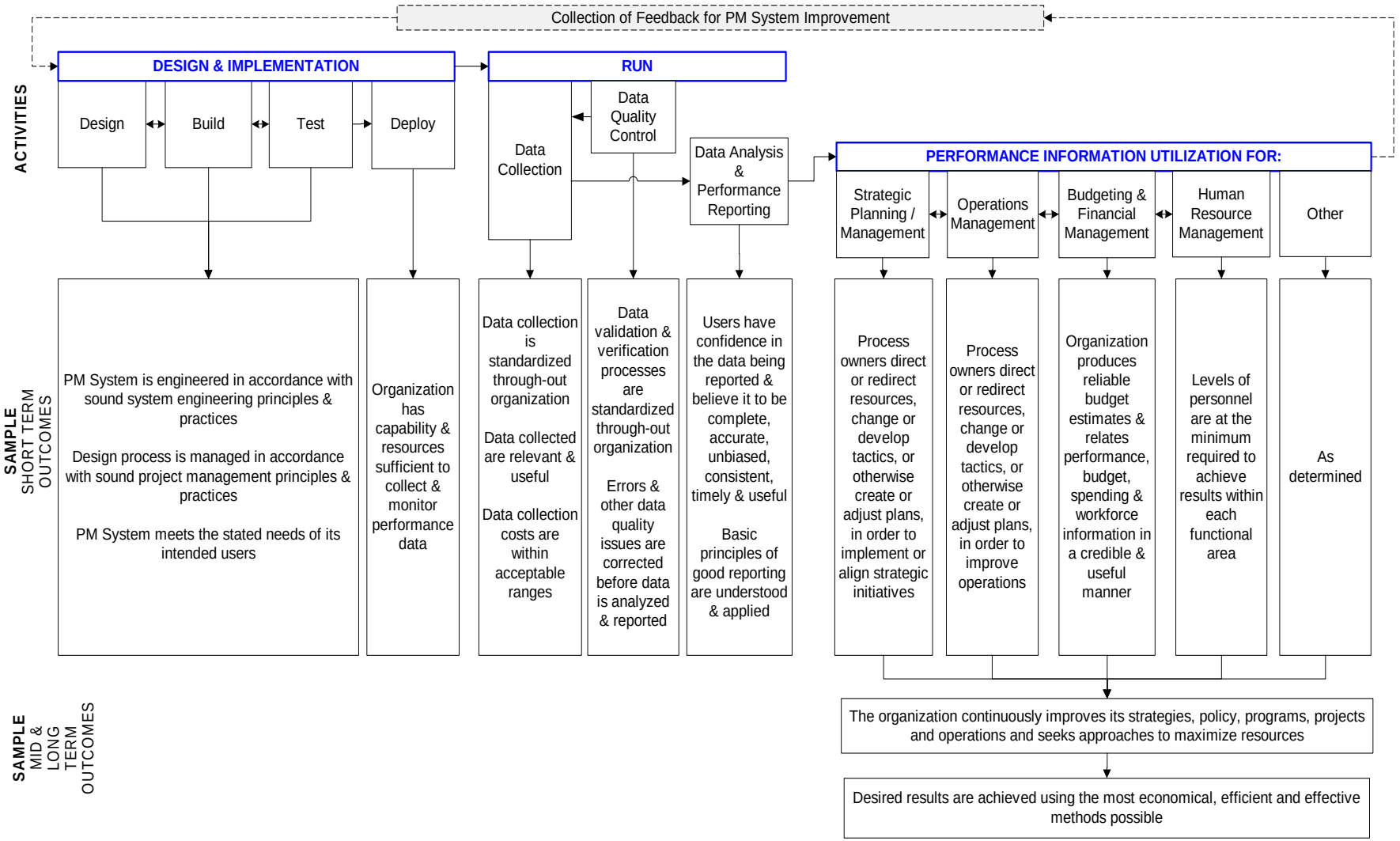

Fig 5: A theory of change for a performance measurement system

PLANNING AND IMPLEMENTING AN EVALUATION USING THE THEORY OF CHANGE

As noted earlier in this paper, a logic model is an attempt to capture the conceptual relationship between long term outcomes and the interventions being implemented so that organizations can demonstrate the probable effectiveness of their activities (Sheirer & Newcomer, 2000; Weiss, 1998). If there are strong logical and/or proven links between the activities and the desired outcomes, there is no need to measure the long term outcomes separately (Weiss, 1998). Further a logic model makes it much easier to organize and focus the evaluation. Thus having a logic model in place at the beginning of an evaluation is a big advantage for the evaluator (Weiss, 1998; Wholey, Hatry & Newcomer, 2004).

Unfortunately, the evaluator has a bit more work to do before the logic model presented in this paper can be used. As mentioned earlier, the first thing that he/she must do is more fully develop the inputs, outputs and outcomes in relation to the performance information utilization activities as these may be specific to particular performance measurement systems. Without a clear understanding of system intentions effectiveness cannot be judged. In cases where system intentions and desired outcomes are vaguely specified (and this paper hypotheses that that will often be the case), the evaluator can question the system's owners until they articulate their assumptions (Weiss, 1998). If this process is not successful then the evaluator should offer hypotheses based on their understanding of the situation (Weiss, 1998), and based on sample outcomes in the logic model presented in Figure 5.

The next steps the evaluator might take will likely need to be done in combination and may iterate several times.

The evaluator in the field should use the logic model in Figure 5 (adapted as appropriate based on the preceding step) to develop research questions that would form the basis of the evaluation plan. This would be accomplished by linking research questions directly to the outcomes in the logic model that are most pertinent to the system under investigation. For example, primary short-term outcomes for *all* implemented performance measurement systems should *always* include those related to the run component, but if the system under study is used only for accountability purposes or only for strategic planning then outcomes specific to the budgetary activity can be safely ignored. Well developed research questions will subsequently suggest appropriate data collection strategies.

At the same time the evaluator would attempt to ensure that performance measurement system activities are supported by academic research or other types of evidence such as promising practice analysis. This step is based on a core evidence-based management principle that says organizations should "be committed to 'fact-based' decision making - which means being committed to getting the best evidence and using it to guide actions" (Pfeffer & Sutton,

2006). Some of the initial work for this step has already been completed through the development of the logic model presented in Figure 5.

Let's work through a specific example to demonstrate how this approach might work. We will use the short term outcome "data collected are relevant and useful" from the data collection activity in the run component of the logic model shown in Figure 5. This is a key outcome that must be investigated carefully. An obvious top level research question that comes to mind is "Are the data collected via the performance measurement system relevant and useful to the system's stakeholders?"

Now we need to define a few terms to better articulate the research question:

- Stakeholders: Who are the system stakeholders? The list of stakeholders should include at a minimum performance measurement system managers, users who collect, record, collate, analyze or report data, users who receive performance reports, as well as any others who may use or benefit from performance measurement system data either directly or indirectly.

- Data collected: What data are collected? This might be readily identified by, for example, examining data collection forms, by reviewing the structure of the system database(s) if any exist, or by reverse engineering existing performance reports.

- Relevant: What do we mean by relevant? (The reader is reminded that what we are talking about here is the relevance of the particular data points or variables that are being collected - not the relevance of the system and not the relevance of the performance reports.) In this case, we can say that the data are relevant to the stakeholder if they can be directly linked to stakeholder needs, which ideally would be articulated through a requirements exercise conducted with stakeholders during the system design activity or after design, articulated through the system's feedback mechanisms.

- Useful: What do we mean by useful? Here we can say that the data are useful if they can be directly linked to a performance report that is generated. If it is not reported, either alone or in combination with other variables, then we can say it is useless.

Now that we have fully defined the research question the evidence collection strategy for it becomes obvious. We need to collect and review data collection forms and/or databases; we need to review system requirements documents and/or other documents that describe stakeholder needs; and we need to review performance measurement system reports. Of course, as the evaluation progresses and other evidence becomes available or if evidence collection strategies fail, then the evaluator would need to adjust the evaluation plan accordingly. Note that the evaluator does not need to determine whether the data collection activity is supported by evidence - this step was already accomplished during development of the logic model illustrated in Figure 5.

A theory of change for a performance measurement system can be used for other purposes as well. For example the evaluator utilizing this model would be in an excellent position to assist in the development of appropriate indicators (e.g., by identifying the outcomes of interest and then seeking evidence of valid indicators as described in the literature) (Sheirer & Newcomer, 2000). The model could also be used as the basis for a process evaluation, a process improvement exercise or as a key component of an evaluability assessment. In fact, each component of a performance measurement system's theory of change - each phase in its life-cycle - presents opportunities for the evaluator to contribute to the ultimate effectiveness of the system. This paper asserts that a theory of change for a performance measurement system (be it the one proposed here or one developed elsewhere) brings with it the same advantages that a good logic model brings to any evaluation (see Weiss, 1998 for an excellent discussion of these advantages).

This paper is enthusiastic in its belief that the proposed logic model will be helpful for the evaluation of performance measurement effectiveness, but it would be remiss not to include at least a brief discussion of issues identified in the literature that seem to influence the effectiveness of performance measurement systems - issues that may not be easily identified in the logic model and that will likely be very challenging for organizations to resolve.

These issues, encountered by organizations in relation to their performance measurement systems, have less to do with the systems themselves and more to do with the environments in which the systems are embedded. For example, Poister (2003) helps us to focus on what should be obvious: that is, some performance problems are just plain difficult to solve and even when good performance information is available and properly fed into decision making processes it can still be ignored or used to make unwise decisions. This is the point at which some authors in the literature seem to conclude that performance measurement systems are ineffective or at best only minimally effective in improving decision making. While we do not necessarily argue with the conclusion drawn, we do argue that evaluating the effectiveness of a performance measurement system using the logic model approach developed here can greatly enhance the evaluator's ability to identify and make recommendations to address these types of problems.

For example, if an evaluator can conclude, with some reasonable level of assurance, that the performance measurement system under investigation is effective, then, by process of elimination, the evaluator might reasonably hypothesize that the problem (e.g., poor performance is ignored or unwise decisions are taken) is occurring or originating outside the measurement system's boundary. Thus the evaluator can refocus investigation into the other management inputs (as illustrated in Figure 1) if the scope of the investigation calls for it.

Furthermore, it might be possible to successfully argue that a performance measurement system could conceivably be designed with these types of seemingly impossible to address problems in mind. The case of 'unwise decisions being taken' is a good example. Unwise decisions made by people within the organization could be identified through the human resource management activity (e.g., during performance reviews) so long as appropriate

indicators were available to guide remedial action (e.g., specific objectives enumerated, number of objectives met).

In any event, further research into these types of issues and their impact on the theory of change presented here seems warranted (see Feller, 2002 for an interesting discussion related to this topic).

CONCLUSION AND IMPLICATIONS

Though little evidence of effectiveness exists, performance measurement systems continue to be implemented based on an assumption that they will have a positive impact on performance (Bourne, Kennerley & Franco-Santos, 2005).

This paper advocates that these systems be systematically evaluated. To address the problem of how best to do this, this paper has proposed a theory of change to guide evaluations. The theory of change presented was developed through careful analysis of the problems and promising practices associated with performance measurement systems as identified in the literature.

The paper offers a methodology for structuring an evaluation of a performance measurement system around this framework.

Using a theory of change to demonstrate effectiveness in circumstances where long term outcomes are expensive or difficult to study is well supported in the field of evaluation (Sheirer &Newcomer, 2000; Weiss, 1998; Wholey, 2002). The premise is that if the key elements of performance measurement system are in place and working as intended (the inputs and outputs are appropriate and short term outcomes are achieved) then one can conclude, with a reasonable degree of confidence, that the performance measurement system is effective. The trick, as this paper has attempted to point out, is to properly identify the key elements of such a system and to define appropriate outputs and outcomes. It may not be easy – an evaluator using this framework must be careful to identify or hypothesize the intended use(s) of the system under investigation before embarking on the evaluation; otherwise, effectiveness cannot be judged.

Evaluators using an evidence-informed theory of change to frame the evaluation of a performance measurement system should be able to assess system effectiveness with a fair degree of confidence. As well, use of the framework and methodology provided here should enable an evaluator to make evidence-informed recommendations for the improvement of a performance measurement system during any stage of the system life-cycle.

A key activity often associated with performance measurement systems is evaluation. That is, performance measurement system data often are intended to inform program, organizational

or policy level evaluations. Thus the evaluation and subsequent improvement of the measurement system itself strengthens the connection between the two.

It is important to note, that while the specific logic model proposed in this paper is intended to be an immediately useful evaluation framework for a performance measurement system, the proposition that a logic model should be used for judging system effectiveness and the methodology used to develop the model may be more useful to the evaluation field in the long run. As systems mature, as improvements are made and as more evidence becomes available, specific activities, inputs, outputs and outcomes may very well shift in focus. No performance measurement system will ever be perfect, but if we begin to evaluate the effectiveness of these systems we can begin to collect a body of evidence that will ultimately either lead to system improvements or to system alternatives.

Because this investigation incorporated literature from a relatively wide range of disciplines and sectors it is hoped that the theory of change presented will be widely applicable and relevant to evaluators in diverse fields.

The information presented in this paper is limited in two ways. First, the model is as yet untested. Case study research is required to test the theory in the field. It is expected that modifications to the theory will occur over time as more evidence becomes available. Second, this paper provides a minimum of sample inputs, outputs and outcomes. More work is required to further develop those parts of the logic model.

Correspondence to: Lori Criss Powers: lpowers@realworldsystems.net

REFERENCES

Auditor General of Canada. (April, 2002). *Report of the Auditor General of Canada to the House of Commons: A model for rating departmental performance reports.* Retrieved 28-March-2009 from http://www.oag-bvg.gc.ca/domino/reports.nsf/html/0206ce.html/$file/0206ce.pdf

Bernstein, D.J. (2001). Local government measurement use to focus on performance and results. *Evaluation and Program Planning*. 24 pp. 95-101

Bititci, U.S., Turner, T. & Begemann, C. (2000). Dynamics of performance measurement systems. *International Journal of Operations & Production Management*. 20(6). P. 692.

Bourne, Kennerley & Franco-Santos. (2005). Managing through measures: a study of impact on performance. *Journal of Manufacturing Technology*. 16(4). pp. 373-395.

Bourne, Mills, Wilcox, Neely & Platts, (2000). Designing, implementing and updating performance measurement systems. *International Journal of Operations & Production Management*. 20(7)

Development Co-operation Agencies. (2000). *Results based management in the Development Co-operation Agencies: a review of experience background report.* Retrieved 28-March-2009from http:// http://www.oecd.org/dataoecd/17/1/1886527.pdf

Ernst, K. (2002). *Hull outcome monitoring and evaluation system - COI outcome model*. Canadian Outcomes Institute. Retrieved 28-March-2009 from http://www.hmrp.net/canadianoutcomesinstitute/pdf/CORIOutcomeModel.pdf

Feller, I. (2002). Performance measurement redux. *American Journal of Evaluation*. 23(4). pp. 435-452.

Franco-Santos, M. & Bourne, M. (2005). An examination of the literature relating to issues affecting how companies manage through measures. *Production Planning and Control*, (16)2. pp. 114-124.

Glaser, B.G. & Strauss, A.L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine De Gruyter.

Hatry, H. & Lampert, L. (2001). *An agenda for action: outcome management in non-profit organizations.* The Urban Institute. Washington D.C.

Henri, J.F. (November, 2004). Performance measurement and organizational effectiveness: bridging the gap. *Managerial Finance.* 30(6) pp. 93-123.

Holloway, J. (2001). Investigating the impact of performance measurement. *International Journal of Business Performance*. 3(2/4). p. 167.

Johnston, R. Brignall, S. & Fitzgerald, L. (2002). 'Good enough' performance measurement: a trade-off between activity and action. *Journal of the Operational Research Society*. (53) pp. 256–262.

Kennerley, M. & Bourne, M. (2003). *Assessing and maximizing the impact of measuring business performance.* Retrieved 28-March-2009 from http://www.som.cranfield.ac.uk/som/research/centres/cbp/downloads/Kennerley%20&%20Bourne.pdf

Kueng, P., Meier, A., & Wettstein T. (2001). Performance measurement systems must be engineered. *Communications of the Association of Information Systems*. 7(3). Retrieved 28-March-2009 from http://citeseer.ist.psu.edu/cache/papers/cs/24632/http:zSzzSzwww2-iiuf.unifr.chzSziszSzpeterzSzKueng_et_al_01.pdf/kueng01performance.pdf

Liner, B., Hatry, HP., Vinson, E., Allen, R., Dunsenbury, P., Bryant, S., Snell, R. (April, 2001). *Making results based state government work.* The Urban Institute. Retrieved 28-March-2009 from http://www.urban.org/UploadedPDF/results-based-stategovt.pdf

Najmi, M., Rigas, J. & Fan, I. (2005). A framework to review performance measurement systems. *Business Process Management Journal*. 11(2). ABI/INFORM Global p. 109

Neely, A.D., Gregory, M.J., Platts, K.W. (1995) "Performance measurement system design – a literature review and research agenda". *International Journal of Operations & Production Management*, 15(4) pp. 80-116.

Perrin B. (1999) Performance measurement: does the reality match the rhetoric? A rejoinder to Bernstein and Winston. *American Journal of Evaluation.* 20(1) pp. 101-111.

Perrin, B. (2003). *Implementing the vision: Addressing challenges to results-focused management and budgeting.* OECD. Retrieved 28-March-2009 from http://www.oecd.org/dataoecd/4/10/2497163.pdf#search=%22Implementing%20the%20vision%3A%20Addressing%20challenges%20to%20results-focused%20management%20and%20budgeting%22

Pfeffer, J. & Sutton, R.I. (2006). What is evidence-based management? Retrieved 28-March-2009 from http://www.evidence-basedmanagement.com/

Plant, T., Agocs, C., Brunet-Jailly, E., Douglas, J. (2005). From measuring to managing performance: recent trends in the development of municipal public sector accountability. *New Directions* #16. International Institute of Public Administration. Canada.

Plant, T. & Douglas, J. (2006). The performance management continuum in municipal government. *Performance Improvement*. 45(1). p. 43

Poister, T.H. (2003). *Measuring Performance in Public and Nonprofit Organizations.* San Francisco, CA: Jossey-Bass.

Poole, D.L., Nelson, J., Carnahan, S., Chepenik, N.G. & Tubiak, C. (2000). Evaluating performance measurement systems in nonprofit agencies: the Program Accountability Quality Scale (PAQS). *American Journal of Evaluation*. 21(1). pp. 15-26.

Reider, R., (2001). *Improving the Economy, Efficiency, and Effectiveness of Not-for-Profits*. New York: John Wiley.

Rivenbark, W.C. & Roenigk, D.J. (2006). A "MAP" approach to performance management in local government. *Government Finance Review*. 22(3) p. 28

Santos, S.P., Belton, V. & Howick, S. (2002). Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations & Production Management.* 22(11). Pp. 1246-1272.

Sheirer, M.A. & Newcomer, K.E. (2000). Opportunities for program evaluators to facilitate performance-based management. *Evaluation and Program Planning*. 24 pp. 63-71.

Smith, P.C. & Goddard, M. (2002). Performance management and operational research: a marriage made in heaven? *Journal of the Operational Research Society*. (53) pp. 247–255.

Tangen, S. (2004). *Evaluation and revision of performance measurement systems.* Royal Institute of Technology, Stockholm.

Treasury Board Secretariat of Canada. (2006). *Results based management*. Retrieved 28-March-2009 from http://www.tbs-sct.gc.ca/rma/rbm-gar_e.asp

United States General Accounting Office (USGAO). (2000). *Managing for results: challenges agencies face in producing credible performance information.* (GAO/GGD-00-52). Retrieved 28-March-2009 from http://www.gao.gov/archive/2000/gg00052.pdf

United States General Accounting Office (USGAO). (2001). *Results-oriented budget practices in federal agencies*. GAO-01-1084SP. Retrieved 28-March-2009 from http://www.gao.gov/new.items/d011084sp.pdf

United Way of America. (2000). *Agency experiences with outcome measurement - Survey Findings.* Alexandria, Virginia. Retrieved 28-March-2009 from http://national.unitedway.org/outcomes/resources/agency_%20experiences.cfm/

Van Aken, E.M., Letens, G., Coleman, G.D., Farris, J. & Van Goubergen, D. (2005). Assessing maturity and effectiveness of enterprise performance measurement systems. *International Journal of Productivity and Performance Management*. 54(5/6). Pp 400-418.

Weiss, C. (1998). *Evaluation* (2nd ed.*)* Englewood Cliffs: Prentice Hall.

Wholey, J.S. (1999). Performance-based management - responding to the challenges. *Public Productivity & Management Review*. 22(3) pp. 288-307.

Wholey, J. S. (1999a). Quality control: Assessing the accuracy and usefulness of performance measurement systems. In H. P. Hatry (Ed.), *Performance measurement: Getting results*. Washington, DC: Urban Institute Press.

Wholey, J.S. (2002). Managing for results: roles for evaluators in a new management era. *American Journal of Evaluation.* (22)3 pp. 343–347.

Wholey, J.S., Hatry, H.P., & Newcomer, K.E. (Eds.). (2004). *Handbook of Practical Program Evaluation* (2nd ed.). San Francisco, CA: John Wiley & Sons.

Wilkins, P. (2004). *The use and usefulness of public sector performance information: have we lost the plot?* Paper to Vision 2020 - Academic Day. Institute of Public Administration, Australia National Conference. Retrieved 28-March-2009 from http://www.griffith.edu.au/school/gbs/ppp/ipaa/wilkins.pdf

Williams, B. & Imam, I. EDs. (2006). *Systems Concepts in Evaluation: An Expert Anthology.* American Evaluation Association. Inverness, CA: EdgePress.

Yourdon, E. (1989). *Modern Structured Analysis*. Upper Saddle River: Prentice Hall.