Evaluation and Program Planning xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Systematic iteration between model and methodology: A proposed approach to evaluating unintended consequences☆

Jonathan A. Morell[a,b]

[a] 4.669 Evaluation and Planning, United States
[b] Director of Evaluation, Syntek Technologies, United States

ARTICLE INFO

Keywords:
Unintended consequences
Evaluation
Complexity
Model
Logic model
Change pattern

ABSTRACT

This article argues that evaluators could better deal with unintended consequences if they improved their methods of *systematically* and *methodically* combining empirical data collection and model building over the life cycle of an evaluation. This process would be helpful because it can increase the timespan from when the need for a change in methodology is first suspected to the time when the new element of the methodology is operational.

The article begins with an explanation of why logic models are so important in evaluation, and why the utility of models is limited if they are not continually revised based on empirical evaluation data. It sets the argument within the larger context of the value and limitations of models in the scientific enterprise.

Following will be a discussion of various issues that are relevant to model development and revision. What is the relevance of complex system behavior for understanding predictable and unpredictable unintended consequences, and the methods needed to deal with them? How might understanding of unintended consequences be improved with an appreciation of generic patterns of change that are independent of any particular program or change effort? What are the social and organizational dynamics that make it rational and adaptive to design programs around single-outcome solutions to multi-dimensional problems? How does cognitive bias affect our ability to identify likely program outcomes? Why is it hard to discern change as a result of programs being embedded in multi-component, continually fluctuating, settings?

The last part of the paper outlines a process for actualizing systematic iteration between model and methodology, and concludes with a set of research questions that speak to how the model/data process can be made efficient and effective.

1. Introduction

This article argues that evaluators could better deal with unintended consequences if they improved their methods of *systematically* and *methodically* combining empirical data collection and model building over the life cycle of an evaluation. This process would be helpful because it can increase the timespan from when the need for a change in methodology is first suspected to the time when the new element of the methodology is operational.

Time matters because evaluation designs are not infinitely flexible. It takes time to succeed in activities such as renegotiating terms of reference, establishing trust relationships with new groups, negotiating for access to new data sources, gearing up data collection in new locations, or identifying new domain experts. Little of this can be done on short notice, yet succeeding in these activities may be needed to evaluate an unintended consequence. Too often we treat unintended consequences as fires. We fight them when they pop up. We do not work methodically at increasing lead time. My belief is that we can do better by deliberately iterating between model building and data collection. We can include planning for unintended consequences along with all the other considerations that go into designing an evaluation (Morell, 2005, 2010).

The article begins with an explanation of why logic models are so important in evaluation, and why the utility of models is limited if they are not continually revised based on empirical evaluation data. It sets the argument within the larger context of the value and limitations of models in the scientific enterprise.

Following will be a discussion of various issues that are relevant to model development and revision. What is the relevance of complex system behavior for understanding predictable and unpredictable unintended consequences, and the methods needed to deal with them? How might understanding unintended consequences be improved with

☆ Website www.jamorell.com, Blog www.evaluationuncertainty.com, YouTube https://www.youtube.com/channel/UCqRIJjhqmy3ngSB1AF9ZKLg.
E-mail address: jamorell@jamorell.com.

http://dx.doi.org/10.1016/j.evalprogplan.2017.09.006
Received 8 September 2017; Accepted 14 September 2017
0149-7189/

an appreciation of generic patterns of change that are independent of any particular program or change effort? What are the social and organizational dynamics that make it rational and adaptive to design programs around single-outcome solutions to multi-dimensional problems? How does cognitive bias affect our ability to identify likely program outcomes? Why is it hard to discern change as a result of programs being embedded in multi-component, continually fluctuating, settings?

The last part of the paper outlines a process for actualizing systematic iteration between model and methodology, and concludes with a set of research questions that speak to how the model/data process can be made efficient and effective. Throughout, following the lead of Koch and Schulpen (2017a, 2017b), I use the term "unintended" to refer to consequences that may be desirable or undesirable, anticipated or unanticipated.

2. Models in evaluation

Models, in the form of program logic models, are dealt with extensively in the field of evaluation (Frechtling, 2007; Knowlton & Phillips, 2007; W. K. Kellogg Foundation, 2004). These well-known sources contain extensive craft knowledge. The sources are rich in examples, instructions on how logic models should be constructed, and guidance on how relevant input can be gained. Missing from these sources, however, is treatment of the epistemology of models. What is the nature of a model? How and when can a model be useful? What are the limitations of models? How can models be wrong, and what are the consequences of that error? Evaluators tend not to ask these kinds of questions and, as a result, miss a great deal of insight into why programs encounter unintended consequences.

Models are needed for the same reason they are needed in all research. They offer a simplified view of reality that distorts and avoids some elements in the service of highlighting relationships that we care about. Without simplification, patterns and relationships that are important would be invisible.

Models may be indispensable, but they can also be fickle. They may be accurate enough for an evaluation's needs over its entire lifecycle, or they can quickly, and often in invisible ways, be overtaken by events. Usually, models do not predict very well (Orrell, 2007). They are not good at prediction because environments change, programs are not stable, stakeholder needs change, factors that affect real-world behavior are inevitably left out, and small events and inaccuracies can result in big changes over time. Further, even when the predictions are acceptable, they will exclude domains of knowledge and understanding that may very well be responsible for unexpected outcomes. This is because, whenever we construct a model, we commit an act of willful ignorance (Weisberg, 2014). To bring forth patterns and relationships that we wish to understand, we deliberately omit others. All in all, George Box's famous aphorism holds: all models are wrong, but some are useful (Box, 1979). In a sense, Box is articulating a very technological view. The model has to further particular purposes. It has to be useful with respect to advancing its intended purpose. It does not have to be true. (For a fuller argument advocating for evaluation as technology rather than science, see (Morell, 1979)).

What, then, makes a model useful? The answer depends on the purpose of the model. If the purpose is to advance *advocacy*, a good model will strip out detail, emphasize the practicality of implementation, and focus attention on the most desirable positive outcomes. These models, however, will not contain enough detail to inform the design of an evaluation that can provide useful information to program planners and funders.

If the purpose of the model is program *design and conduct*, the model takes the form of a project or process plan. This plan will describe what steps need to be taken when. But operational detail does not translate into understanding of how steps in running a program translate into impact, or what it is about the program that can be expected to result in particular outcomes.

If the model's purpose is to *explain* why a program works as it does, the model will contain a great deal of specificity about program elements and outcomes, and the relationships among them. These models, however, are likely to be far too elaborate to drive data collection and analysis.

Finally, there is the purpose of models that concerns me in this paper, namely, as a guide to designing evaluation methodology and data collection schemes. Or put another way, my concern is with models that depict hypotheses about program outcome that can be tested empirically. The models have to help structure evaluations that can be completed within a reasonable time period, executed with a reasonable budget, and require the collection of data that can be collected in the setting where the evaluation is to take place.

In addition to practicality, other characteristics of models are also desirable.

1. The model should convey a qualitative sense of change, e.g.
   - Is change continuous or discontinuous, sudden or gradual?
   - Will outcomes appear immediately after a program is implemented, or only after some time has elapsed?
2. The model should not mix levels of detail without good reason.
   - As an example, imagine a program to improve cooperation between one group of programs aimed at improving agricultural production, and another designed to support export growth. The theory is that, by coordinating formal meetings among the groups, the richness of cooperation among them would increase.
   - Should the evaluation include a measure of attitude change that came about as a result of the *particular way* the meetings were conducted? Including that measure would imply a belief that a particular method of running the meetings was necessary for facilitating cooperation. Do we believe that? Even if we did, does it matter? If the answers to these questions is "yes," then we should include the measure and the methodology it would require. If not, why work with a model that implicitly assumes the importance of that lower-level theory, while adding cost and complexity to the evaluation?
3. The model should be honest with respect to what we do and do not know about the program, e.g.
   - Do we really believe that the single causal path that we have traced among outcomes is the only one that may occur?
   - Do we really believe that all those relationships that we have specified are actually how things work?
   - Are we sure that those intermediate outcomes in the model are really needed for the program to produce the results we want it to?
   - When it comes to model building, do we fool ourselves into believing that we know more than we really do?

These desirable characteristics of models are agnostic as to form. The models can be qualitative or quantitative. They can have rich or sparse detail. They can have short- or long time horizons. They can contain one or multiple levels of detail. They can specify any number of relationship types: 1:1, 1:many, many:1, or many:many. Or, they might make do with collections of elements and omit detail as to relationships among elements within the collection. The form of the model is irrelevant. It may be as simple as a few lists of words in columns, or extremely precise and elaborate. If the model can guide good program evaluation, that's all that counts.

What does it take to guide good program evaluation? And in particular, what does it take to guide evaluation that is sensitive to unintended consequences? One part of the answer goes beyond models as artifacts, and extends to the process by which models are developed and used. Another part of the answer is assuring that model building accounts for the kinds of changes that generate unintended consequences. Process is critical. Without it, models cannot help form methodology