Joining Minds:
Group Modeling to Create Public Value

George P. Richardson*  
gprichardson@albany.edu  
Rockefeller College of Public Affairs and Policy  
University at Albany  
Albany, NY 12222

David F. Andersen  
david.andersen@albany.edu  
Rockefeller College of Public Affairs and Policy  
University at Albany  
Albany, NY 12222

Luis F. Luna-Reyes  
luisf.luna@udlap.mx  
Universidad de las Américas-Puebla  
Business School, CS213 F  
Sta. Catarina Mártir, Cholula, Puebla, MEXICO  
72820

*Contact author

George P. Richardson is O’Leary Professor of Public Administration and Policy and Information Science in the Rockefeller College of Public Affairs and Policy at the University at Albany. He has a PhD in Management from the Sloan School at MIT. He is the author of more than thirty articles on the theory and application of system dynamics modeling to problems of group decision support in the public sector, as well as Introduction to System Dynamics Modeling with DYNAMO (1981), Feedback Thought in Social Science and Systems Theory (1991) and the edited collection Modeling for Management: Simulation in Support of Systems Thinking (1996).

David F. Andersen is Distinguished Service Professor of Public Administration and Information Science at the Rockefeller College of Public Affairs and Policy, University at Albany. He holds a Ph.D. in Management from the MIT Sloan School of Management. His recent work has focused on the use of formal computer-based tools and models to help groups understand the system-wide impacts of information and decision support systems. Professor Andersen's consulting and research activities include assembling and managing research and consulting teams to address a wide variety of public policy problems.

Luis Felipe Luna-Reyes is an Associate Professor of Business at the Universidad de las Américas in México. He has a PhD in Information Science with a primary focus on Group Decision Support Systems for the University at Albany. His research focuses on collaborative information systems development across functional and organizational boundaries, dynamic modeling of socio-technical systems, and electronic government.
Joining Minds:

Group Modeling to Create Public Value

Abstract

System dynamics group model building is a process that engages public managers and policy makers in strategic discussions that can create Public Value for their organizations, their direct clients, and the public at large. These techniques link people, process, analysis, and policy design as a strategy to create Public Value. System dynamics group model building involves key stakeholders, group facilitation, and formal computer simulations in a transparent process designed to join minds. The goal is to assist public managers and policy makers grappling with a complex problem to achieve a policy consensus. This article links more than fifteen years of experiences with system dynamics group modeling as documented in the systems thinking and management sciences literatures, to the public management literature, emphasizing the potential of the approach to generate Public Value. It outlines the process in an extended example dealing with welfare reform, and explores evaluation studies to identify conditions that affect the success or failure of these initiatives. The appeal of system dynamics group modeling to public managers is traced to its abilities to engage stakeholders, build on their mental models to help them define public value, and building operational capacity to handle extreme complexity, facilitate alignment, enable refutable “what-if” thinking, and empower the group—all of these key strategic elements in creating Public Value.
What is Group Model Building?

In 1987, New York State was facing a crisis in its medical malpractice system. Doctor and hospital-paid premiums into the state-sanctioned system were skyrocketing. Physicians, especially obstetricians, were refusing to treat new patients and some were moving out of the state to practice elsewhere. The governor and the legislature were caught in crossfire between hospital associations, medical associations, trial lawyer associations and insurance carriers. Pressing for dramatic hikes in doctors’ premiums, actuarial calculations predicted that the malpractice insurers’ reserve funds were statistically insolvent to the tune of $2 billion. A problem of this magnitude, based on complex legal and technical arguments and involving the core interests of competing powerful stakeholders, would now be characterized in the literature as a “messy” or “wicked”\(^1\) problem.

As part of a solution to this crisis, the Commissioner of the New York State Insurance Department convened a group of expert political, financial, and actuarial stakeholders internal to his agency to contemplate and design possible solutions to this connected set of issues. As previously reported in the management science literature (Reagan-Cirincione et al. 1991), with the support of the Decision Techtronics Group (DTG),\(^{2}\) the collected group designed and implemented a series of computer-based simulation models and multi-criteria decision-making models that helped to guide the state toward a resolution of the crisis that focused squarely on enhancing Public Value.

These meetings became the first published example of what systems dynamics modelers later came to call Group Model Building (GMB). For sake of clarity here, we shall refer to this process as System Dynamics Group Model Building (SDGMB) to distinguish it from other
forms of group-oriented modeling practices. Four developments had come together right around 1988 to make the first case of SDGMB possible. First, the field of computer simulation and modeling was maturing and moving toward working directly with multiple competing stakeholder groups so characteristic of the public sector. Second, micro-computing and early computer projection technology had made it possible for the first time to bring live computer support into group meetings. Third, methods of group facilitation, long a subject of study for small group experts, had formed the basis for the newly emerging field of Group Decision Support Systems (GDSS). Fourth and finally, the recent development of icon-oriented system dynamics simulation software (initially STELLA, later Vensim and Powersim Studio) enabled modelers to develop simulation models at a high level in real time with live client groups in the room. While many of the formulation and calibration tasks of the modeler still remained “in the back room,” key system conceptualization and problem defining tasks could now be accomplished in front of, and with the active involvement of, client groups representing diverse stakeholder interests.

Since 1987, the field of SDGMB has made considerable progress. It has diffused out from its origins in public sector work (where it was uniquely suited to support complex and conflicting stakeholder groups of clients) to become an emerging consulting practice used in private, public, and not-for-profit organizations. In a recent meta-analytic survey, Rouwette et al. (2002) reported on over one hundred cases appearing in the recent published literature. A literature documenting systematic methods of carrying out SDGMB has emerged, giving practice guidelines to consultants and practitioners new to the field. Finally, researchers are beginning to grapple with the complex issues of defining and measuring the effectiveness of this class of interventions.
This paper presents an update on what has been happening with SDGMB since its inception in 1987 and links that ongoing work to efforts to generate Public Value. Table 1 provides thumbnail sketches of eleven illustrative interventions from our group, tracing some of the recent developments in the art and practice of group modeling. After presenting an interpretation on Public Value that is embodied in these studies, we present an extended example. Then we briefly review some of the emerging literature on methods, overview several of the key studies that are evaluating group modeling efforts, and conclude with reflections on current practice and future prospects.

**Table 1: Thumbnail sketches of a number of published group model building interventions with a public sector focus.**

<table>
<thead>
<tr>
<th>Client Organization</th>
<th>Problem</th>
<th>Processes</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYS Department of Social Services</td>
<td>Understanding foster care caseload dynamics</td>
<td>Group model building workshops with foster care experts, a model-based masters essay with policy analyses</td>
<td>Continuation of research in NYC and NYS DSS and Chicago on the dynamics of foster care caseloads</td>
</tr>
<tr>
<td>[Ref 1]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermont Department of Health and Human Services</td>
<td>Rising Vermont Medicaid costs</td>
<td>Group model building with VT Medicaid experts and stakeholders</td>
<td>Resistance from traditional quantitative approaches; no clear implementation of system dynamics insights</td>
</tr>
<tr>
<td>[Ref 2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NY City Office of Management and Budget</td>
<td>Caseload growth in NYC foster case</td>
<td>Group model building workshop with practitioners and NY City managers and OMB</td>
<td>No clear policy implementations. Motivation for continued research between NYC practitioners and NYS DSS researchers</td>
</tr>
<tr>
<td>NYS Office of Mental Health</td>
<td>Studying the failed efforts to integrate vocational services of OMH and VESID for mental health clients</td>
<td>Group model building workshops in four counties, building shared understandings between OMH and VESID county workers</td>
<td>Comparisons with four similar counties without the systems workshops showed the group modeling increased alignment, achieved</td>
</tr>
<tr>
<td>[Ref 3]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Description</td>
<td>Methods</td>
<td>Accomplishments</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>---------</td>
<td>-----------------</td>
</tr>
<tr>
<td>NYS Office of Real Property Services [Ref 4]</td>
<td>Moving from an adversarial role for ORPS with localities to a consulting model</td>
<td>Several group modeling workshops involving both system dynamics and strategic mapping</td>
<td>Increasing clarity in the nature of a consulting role for ORPS with localities, and enhanced motivation and capabilities to pursue it</td>
</tr>
<tr>
<td>National Cancer Institute [Ref 5]</td>
<td>Dynamics of tobacco prevalence and control, as a test of ISIS, the Initiative for the Study and Implementation of Systems in public health</td>
<td>Policy-oriented systems maps and models of the dynamics of prevalence and control, linked with the other ISIS approaches: networks, knowledge management, and organization change</td>
<td>Research report capturing the individual and combined contributions of the four components of the ISIS approach to studying tobacco prevalence and control efforts.</td>
</tr>
<tr>
<td>Center for Technology in Government (at UAlbany) and NSF [Ref 6]</td>
<td>Building grounded theory on implementing information integration among public sector agencies</td>
<td>Several group model building workshops with CTG staff evolving a model representation of the structure and dynamics of their interventions</td>
<td>Increasingly sophisticated model-based understandings (theories) of the problems and possibilities involved in implementing information integration among public sector agencies</td>
</tr>
<tr>
<td>Consortium of off-shore oil platform developers interested in developing wind farms [Ref 7]</td>
<td>Financial feasibility of off-shore wind farms depended on unknown mix of risk and govt policy factors</td>
<td>Consortium used group model building to understand the structure and dynamics of the off-shore wind farm market</td>
<td>Simulations identified and quantified risk factors; report was sent to national legislative bodies.</td>
</tr>
<tr>
<td>British National Health Services [Ref 8]</td>
<td>Provision of long term dementia services in the Scottish Borders was straining service capacity</td>
<td>Local coalition of service providers met in group mapping sessions to plot new directions</td>
<td>Sessions did not result in a formal running simulation, but did lead to a service integration proposal</td>
</tr>
<tr>
<td>CERT National Cyber-Security Center</td>
<td>Insider Threats provided an ill-understood source of cyber vulnerability for the USA internet</td>
<td>Cyber-security experts joined behavior scientists to craft a dynamic theory of cyber</td>
<td>Published theory of insider attacks supports national policies on cyber-security aimed at</td>
</tr>
</tbody>
</table>
Public Value

Although the concept of public value is not new, current discussions have been invigorated by Moore’s work in 1995 (Benington and Moore, 2011). The main objective of the book was to provide public managers with a framework to help them understand their strategic alternatives in the process of creating public value. Three key concepts of public value introduced in the book were the strategic triangle, the authorizing environment and the use of state authority. The strategic triangle is a framework showing the importance of aligning three interrelated processes: defining Public Value, building and sustaining a group of diverse stakeholders to create an authorizing environment, and mobilizing the resources from inside and outside the organization to achieve the desired outcomes (Benington and Moore, 2011).

Each of the group model building interventions shown in Table 1 had its foundation in an effort to generate Public Value highly consistent with these three key concepts. Additionally, we build
on the extensive work in this area of the Center for Technology in Government at the University at Albany [See http://www.ctg.albany.edu/projects/opengov]. CTG traces the beginnings of their work to the Open Government Directive of the Obama administration. They extend the three dimensions emphasized in that directive – transparency, participation, and collaboration – to the six generators of Public Value shown in Table 2.

Table 2: Generators of Public Value [Source: CTG]

<table>
<thead>
<tr>
<th>Generator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>obtaining increased outputs or goal attainment with the same resources, or obtaining the same outputs or goals with lower resource consumption.</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>increasing the quality of the desired outcome.</td>
</tr>
<tr>
<td>Intrinsic enhancements</td>
<td>changing the environment or circumstances of a stakeholder in ways that are valued for their own sake.</td>
</tr>
<tr>
<td>Transparency</td>
<td>access to information about the actions of government officials or operation of government programs that enhances accountability or influence on government.</td>
</tr>
<tr>
<td>Participation</td>
<td>frequency and intensity of direct involvement in decision making about or operation of government programs or in selection of or actions of officials.</td>
</tr>
<tr>
<td>Collaboration</td>
<td>frequency or duration of activities in which more than one set of stakeholders share responsibility or authority for decisions about operation, policies, or actions of government.</td>
</tr>
</tbody>
</table>

These generators are not ends in themselves, but rather “instrumental to the accomplishment of democracy,” enabling “citizens to perform their roles as citizens” (op. cit.) The work of CTG further emphasizes the multiple and diverse stakeholders that would be involved in any in-depth analysis of an initiative designed to create Public Value.

System dynamics group modeling targets the three traditional factors in the strategic triangle, and all six of the generators of Public Value shown in Table 2 and does so in contexts made complex not only by the difficulty of the problems but also by the diverse stakeholders involved.
In order to give a better sense of what happens in one of these group model building projects and how they target Public Value, we describe in the next section an eighth example that used SDGMB to support welfare reform initiatives in three county governments in New York State. The section is based on a note published in the *Journal of Policy Analysis and Management* (Zagonel et al. 2004). We chose to present this example because it is particularly well-documented, with aspects of the project having been reported previously in the literatures on public affairs (ESR 1998; Rohrbaugh and Johnson 1998; Rohrbaugh 2000) and simulation (Rogers et al. 1997; Allers et al. 1998; Lee, Zagonel et al. 1998; Lee 2001; Andersen et al. 2000; Zagonel 2003).

**Background of the case**

In 1996, then President Clinton signed into law the Personal Responsibility and Work Opportunity Reconciliation Act. While many of the implications of this law have been made clear with the passage of time (Ewalt and Jennings 2004), at its inception policymakers and researchers had no steady intuition to guide them in considering plausible outcomes: What if there was a recession while clients are facing time-limits? How can counties cope with increased demands for scarce resources as clients lose Federal welfare eligibility? What if neighboring states are forced to cut back benefits? In New York State, these questions have a special edge since Article 17 of the state constitution mandates local governments to provide for the indigent and needy. While the federal government can end entitlements after five years, New York State and its local governments cannot suspend all benefits.
A coalition of state agencies and county governments set out to use simulation technologies supported by SDGMB techniques to address these questions (Richardson and Andersen 1995; Andersen and Richardson 1997; Vennix 1996; Vennix, Andersen, and Richardson 1997).

A Thumbnail Sketch of the New York State Welfare Simulation Project

The project emerged in four overlapping phases. The first phase involved building a simulation model of the basic Temporary Assistance to Needy Families (TANF) system for Cortland County, a rural county located in central New York State (Rogers et al. 1997; Rohrbaugh 2000; Zagonel 2003). In this phase a preliminary simulation model was constructed with the Commissioner and her management team using SDGMB.

The second phase of the project concentrated on formulating the “Safety Net” sectors of the model, which would serve clients after losing eligibility as mandated by New York State constitution. This portion of the modeling effort imported the TANF model developed in Cortland County, but had it calibrated and joined with the safety net model developed for Dutchess county, a mid-sized suburban county located in the Hudson valley (Allers et al. 1998; Zagonel 2003). The simulation model was extended and elaborated.

The third phase of the project was conducted in conjunction with a network of service providers located in Nassau County, a large and demographically complex county directly adjacent to New York City. After calibrating the model for this area, the group explored how approaches developed in smaller regions applied in the more complex environment in the metropolitan area, adding yet more refined detail to the simulation model.
The final phase of the project was aimed at implementing policy insights from the model.

Follow-up workshops were held in these counties with broadly-based groups of stakeholders, including new participants who were not involved in model development. To facilitate this process, the joined TANF-Safety Net model was wrapped in a graphical user interface that allowed it to be used by managers, policy makers, and laypersons who had not been previously involved in the model construction process (ESR 1998; Rohrbaugh and Johnson 1998).

**Policy Insights from the Simulation Model Contributing to Public Value**

An important policy insight produced by experiments through the modeling process involved the comparison of two investment policies labeled “Edges” and “Middle.” The Base run (or “reference” run) shows the model’s projection of what would happen to the county’s welfare system if no policy changes were made and if there are no external scenario changes. The “Middle” policy simulated a high investment in assessment, monitoring, and job-finding and promotion functions traditionally associated with a social services unit. The “Edges” policy contained a mixture of resource investments that concentrated on the “front” and “back” ends of the system (i.e. Prevention, Child Support Enforcement, and Self-Sufficiency Promotion).

Figure 1 compares the Base run to the Middle and Edges policy packages for one key performance indicator—total **job finding flows** from TANF. The X-axis is time, measured in years, and the Y-axis is the total flow of people out of TANF, measured in people per year. As shown in the figure, significant new investment in the Middle policy package greatly accelerates this presumably beneficial trend. Investing in the Edges of the system appears in the simulation actually to retard Job-Finding relative to the Base run.
Figure 1: Total Job-Finding Flows from TANF (Base vs. Middle vs. Edges Policy Packages) —The base case provides the initial reference point (2400 people/year in 1997) matching history in the particular county, against which to compare the alternative policy scenarios. But the conclusion from Figure 1 is misleading. By focusing on a less commonly articulated performance measure —total recidivism— Figure 2 reveals an important structural insight lurking behind these graphs. The Edges policy has the effect of significantly reducing recidivism in the model. Since the total number of persons coming on to TANF at any point in time is the sum of first-time recipients plus recidivists, this reduction in recidivism can dramatically decrease the overall TANF caseload. By contrast, the Middle policy actually has the simulated effect of increasing recidivism. Richardson et al. (2002) have demonstrated that in the simulation model the high influx of families on TANF into the post-TANF employment support system had the effect of “swamping” these downstream resources, leading to long-term increases in recidivism. Since recidivism is not mandated in the Federal legislation to be tracked and is hard to document across the welfare system, the increased TANF caseload could easily be misinterpreted as the result of some external influence such as rising unemployment rather than as a natural, endogenous consequence of the Middle policy intended to reduce caseloads.
To summarize the mechanism at work here, the Middle policy is great at getting people into jobs, but then they lose those jobs and cycle back into the system because there aren’t enough resources devoted to help them stay employed. The Edges policy lets them trickle more slowly into jobs but then does a better job of keeping them there.

Whatever the final policy choice, the simulation model provides a “level playing field” for evaluating the implications of multiple policy and scenario changes always using precisely the same agreed-upon set of assumptions and numbers. Of course, these inferences are only as good as the model upon which they are based. But here, SDGMB has the great strength that the model and its parameters emerge from facilitated conversations among experts in the problem. In addition, the overall project paid careful attention to qualitative and quantitative model testing, sensitivity analyses, and calibration (Lee, Andersen et al. 1998b; Lee, Andersen et al. 1998a; Lee, Zagonel et al. 1998; Zagonel et al. 2004).
This one aspect of the group model building effort with NY State counties focusing on the significance of the structure and dynamics of recidivism illustrates the contributions of the approach to Public Value. The group introduce the idea of recidivism as a measure of public value, and the analyses were transparent: model building and model evaluation took place in public. The work was participatory and collaborative: the model was not the work of one group but rather a highly participative process with most stakeholders present and active. And the diverse group was actively involved in stressing the model to test it and to help build confidence in its structure and dynamic implications.

**What happened in the Three Counties?**

The pattern of pursuing implemented results varied considerably in the three counties. The Commissioner of Social Services in Cortland County used the model and its results to plan her investment priorities to implement Welfare Reform for the mid- to long-term (Rogers et al. 1997). In Dutchess County, Commissioner Allers created a public-private task force to design and implement concrete initiatives aimed at the “Edges” policies (Allers et al. 1998; Rohrbaugh and Johnson 1998; ESR 1998; Rohrbaugh 2000). Implementations in Nassau County were the least extensive of the three sites. The Commissioner convened her direct staff to work with the model but did not involve a wider group of community stakeholders to implement model-based implications.

**How The SDGMB Method Works**

There is much to say about approaches to group conversations about messy problems. There are many approaches all presumably designed to enhance Public Value. Below we describe the
defining characteristics of the SDGMB approach and touch upon some practical considerations associated with helping groups of managers develop policy and strategy.

Team Facilitation

Experts in group decision support point to skilled facilitation as a key success factor to any group method (Bryson and Anderson 2000). Because of the nature of the tasks involved in facilitating SDGMB sessions and model development, it is complicated for a single person to accomplish all of them. SDGMB has been characterized by the use of team facilitation. Each member of the facilitation team plays a different role in a coordinated manner. Through the experience along the years, our group has identified five different roles: facilitator, modeler/reflector, process coach, recorder and gatekeeper. Richardson and Andersen (1995) have described these five roles in detail. A key feature of creating and sustaining multiple roles is the ability to separate the facilitation work into two specialized tasks; facilitation and analysis (Rohrbaugh 1992; Zagonel 2002; Vennix et al. 1994; Richardson and Andersen 1995).

Scripts for Group Model Building

SDGMB sessions are the result of a very careful planning process (Andersen and Richardson 1997). Tasks are usually broken up into discrete periods of 15 to 20 minutes duration throughout the day, promoting continuous change of activity and group dynamics to improve group engagement. A variety of convergent, divergent, and evaluative group tasks are cast as “scripts” all of which cumulate to create pieces of model structure that can be assembled into the final formal simulation model. Usually, the client group sits in a U-shape form, with the facilitator in the front with a computer or an overhead projector. The ideal room is that one with lots of place
to write down diagrams and ideas. The modeler/reflector, and the recorder sit at the back of the room. Figure 3 shows a room setup in a picture taken from the recorder location.

Figure 3: Room setup in a typical SDGMB session

Detailed descriptions of the process and scripts used in SDGMB have been reported in the simulation literature (Andersen and Richardson 1997; Andersen et al. 2004; Luna-Reyes et al. 2006). Although the planning stage is very important, the facilitation process requires flexible improvisation after compulsively detailed advance planning (Andersen and Richardson 1997).

**In-meeting and Between-meeting Work of Team and Clients**

A typical SDGMB project involves effort from the facilitation team and the client team both during the SDGMB sessions, and between them. The main activities of the facilitation team include facilitating the sessions, working in model development, coordinating the effort, writing reports, and gate keeping (Luna-Reyes et al. 2006). Besides participating in the SDGMB
sessions, the client group participates in defining the scope of the project and in providing soft and hard data for model calibration. At the present state of the art, much of the technical model formulation, testing, and validation are completed by the SDGMB team outside the group sessions.

**Recent Developments in SDGMB: From Artful Practice to Scripted Craft**

One of the limitations to the use of SDGMB to support the generation of public value has been the perception, if not reality, that actually completing a group model building project was an art form that blended multiple talents that are hard to assemble in a single team. Typically a SDGMB team would involve a skilled system dynamics modeler as well as at least one more person with facilitation skills. Often a recorder joined the modeling team. In addition to needing to assemble all of these specific skills, a team needed to have training in how to organize a full day or more of activities with a client group. While the benefits of these sessions in terms of generating public value are high, the barriers to doing successful work were also high and the risks of a less than successful session with high client involvement could prove to be a barrier to successful implementation.

Recently Hovmand et al. (2012) have been working to make SDGMB less of an art form and more of a structured craft. They have published a series of scripts that describe precisely small bits of behavior that a skilled facilitator or modeler should engage in when working with a group. The intention of this “Scriptapedia” is to produce an easy-to-use handbook that would allow teams with little background in SDGMB to get started. Increasingly, teams with no experience in using SDGMB in a public value creation project have been able to pick up the manual and begin to make progress working with a group.
Ackermann et al. (2010) have recently contributing to this conversation by creating what they call a “Scriptsmap” or a structured system for “snapping together” various smaller scripts into a larger all day workshop for the purpose of working with client groups in a public value creating exercise. These two recent developments promise to move the art of SDGMB into a more easily replicable form of craft.

**Evaluating SDGMBs**

SDGMB is a promising technology for creating Public Value. But does it really work? Over the past several decades, a minor cottage industry has emerged that purports to document the successes (and a few failures) of SDGMB by reporting on case studies. Telling stories much the same as the ones we have presented in this paper, authors of these case studies allege to report on successful applications and sometimes analyze weaknesses, making suggestions for improvement in future practice. Indeed, Rouwette et al. (2002) have compiled a meta-analysis of 107 such case-based stories.

However compelling such case stories may be, case studies are a famously biased and unsystematic way to evaluate effectiveness. Presumably, failed cases will not be commonly reported in the literature. In addition, such a research approach illustrates in almost textbook fashion the full litany of both internal and external threats to validity, making such cases an interesting but unscientific compilation of war stories. Attempts to study live management teams in naturally occurring decision situations can have high external validity but almost always lack internal controls necessary to create scientifically sound insights.
Hsiao (1999) added a new perspective to this research by reviewing 34 laboratory studies of how human subjects acting as surrogate managers and policy makers learn from complex computer simulation models. In these experiments, individuals (not groups) interact with simulated policy micro-worlds (not messy policy realities). Even in these tightly controlled circumstances, the evaluation literature struggles to articulate how human beings learn to grapple with and control dynamic complexity. This is complicated stuff. But Hsiao reports that at least in controlled settings we are starting to learn how better decisions can and do emerge as human actors learn about complex policy dynamics from simulation models. But what is known about managers working in the field?

McCartt and Rohrbaugh (1995) returned to 26 live decision conferences conducted mostly in the public sector by the Decision Techtronics Group (DTG) at the University at Albany. Using a survey methodology, they polled all of the participants of these decision-making conferences one to three years after the conferences had taken place. They were searching for lasting mid- to long-range effects of these decision conferences as well as a deeper understanding of what factors caused such conferences to be successful (or not). They found that the introduction of group decision support technologies substantially changes the way that groups function and how decisions are arrived at. Approximately one fourth of the client organizations surveyed had results that they characterized as “low success”. When seeking to explain differences between low and high success conferences, they found that

GDSS adoption is more likely in client organizations that are open and receptive to change, in executive teams that place considerable value on adaptation, and in management groups that attach great importance to flexible and creative decision processes. Evidence also indicates that the introduction of GDSS can result in important decisions that prove beneficial over time when the organizational functions of goal attainment and integration are emphasized (assuring that decision processes are viewed by group members as providing considerable efficiency and accountability). Finally, our findings appear to favor applications of GDSS that promote full (but
not necessarily equal) participation, that encourage raising questions and expressing concerns, and that facilitate the constructive use of conflicting opinion (McCartt and Rohrbaugh 1995, 581).

Working with the same data set as McCartt and Rohrbaugh (1995), Schuman (1995) concluded that “Important decisions are more likely to be made when there is a strong empirical decision-making process.” However, “important decisions are more likely to be made when participants place little emphasis on the importance of data.” This apparent contradiction suggests that an effective empirical process must rely on structural or other verbal information that is not in the form of data captured in a spreadsheet or database.

Huz et al. (1997) created an experimental design to test for the effectiveness of SDGMB. They noted that “the multiple treatments, pre-test/post test control-group design provides some controls for many threats to internal and external validity” (156). They used a wide battery of pre and post survey, interview, archival, administrative data, and qualitative observation techniques to evaluate eight carefully matched interventions. All eight interventions dealt with the integration of mental health and vocational rehabilitation services at the county level. Four of the eight interventions contained SDGMB sessions and four did not. Four of the eight counties received an explicit mandate to integrate services (linked to a series of waivers of regulation and relaxation of categorical funding streams) and four did not. All-in-all, the eight counties made a 2x2 treatment grid with two counties in each cell.

Overall, Huz et al. (1997) envisioned that change could take place in nine domains spread measured across three separate levels of analysis as illustrated in table 2 below:
Table 3: Domains of measurement and evaluation used to assess impact of systems-thinking (SDGMB) intervention (Huz et al. 1997,151)

<table>
<thead>
<tr>
<th>Level I</th>
<th>Reflections of the modeling team</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 1</td>
<td>Modeling team’s assessment of the intervention</td>
</tr>
<tr>
<td>Level II</td>
<td>Participant self-reports of the intervention</td>
</tr>
<tr>
<td>Domain 2</td>
<td>Participants’ perceptions of the intervention</td>
</tr>
<tr>
<td>Domain 3</td>
<td>Shifts in participants’ goal structures</td>
</tr>
<tr>
<td>Domain 4</td>
<td>Shifts in participants’ change strategies</td>
</tr>
<tr>
<td>Domain 5</td>
<td>Alignment of participant mental models</td>
</tr>
<tr>
<td>Domain 6</td>
<td>Shifts in understanding how the system functions</td>
</tr>
<tr>
<td>Level III</td>
<td>Measurable system change and “bottom line” results</td>
</tr>
<tr>
<td>Domain 7</td>
<td>Shifts in network of agencies that support services integration</td>
</tr>
<tr>
<td>Domain 8</td>
<td>Changes in system-wide policies and procedures</td>
</tr>
<tr>
<td>Domain 9</td>
<td>Changes in outcomes for clients</td>
</tr>
</tbody>
</table>

Using the battery of pre- and post-test instruments, Huz et al. (1997) were able to find important and statistically significant results in eight of the nine domains measured. The exceptions were in domain 9 where they did not measure client outcomes, in domain 5 where “participants were not significantly more aligned in their perceptions on strategies for changes” (but were more aligned in goals), and in domain 7 where “no significant change was found with respect to structural conditions within the network” (but two other dimensions of organizational relationships did change).

In their meta-analysis of 107 case studies of SDGMB, Rouwette et al. (2002) coded case studies with respect to eleven classes of outcomes, sorted into individual level, group level, and organizational level. The 107 cases were dominated by for-profit examples with 65 such cases
appearing in the literature followed by 21 cases in the non-profit sector, 18 cases in governmental settings, and three cases in mixed settings. While recognizing possible high levels of bias in reported cases as well as difficulties in coding across cases and a high number of missing categories, they found high percentages of positive outcomes along all 11 dimensions of analysis. For each separate dimension, they analyzed between 13 and 101 cases with the fraction of positive outcomes for each dimension ranging from a low of 83% to several dimensions where 100% of the cases reporting on a dimension found positive results. At the individual level, they coded for overall positive reactions to the work, insight gained from the work, and some level of individual commitment to the results emerging from the study. At the group level, they coded for increased levels of communication, the emergence of shared language, and increases in consensus or mental model alignment. Organizational level outcomes included implementation of system level change. With respect to this important overall indicator they “found 84 projects focused on implementation, which suggests that in half (42) of the relevant cases changes are implemented. More than half (24) of these changes led to positive results” (Rouwette, Vennix, and Van Mullekom 2002,20).

Rouwette (2003) followed this meta-analysis with a detailed statistical analysis of a series of SDGMB interventions held mostly in governmental settings in the Netherlands. He was able to estimate a statistical model that demonstrated how SDGMB sessions moved both individuals and groups from beliefs to intentions to act and ultimately on to behavioral change.

Most recently, Zagonel (2004) drilled down to examine the micro-level scripts being used in seven separate records of a single, prolonged SDGMB project. His purpose was to document tensions in the SDGMB process. On the one hand, he viewed models as “boundary objects” that
can be used to negotiate policy directions with client teams even if they are separated by various competing points of view, in effect using the group modeling process to bridge the boundaries that separate them. On the other hand, models can be seen as “micro-worlds” that can be used as data-based representations of a policy reality, helping client teams to see the over-time implications of their decisions. Zagonel’s work was designed less to assess outcomes for the clients and more to help design better SDGMB interventions.

In sum, attempts to evaluate SDGMB interventions in live settings continue to be plagued by methodological problems that researchers have struggled to overcome with a number of innovative designs. What is emerging from this body of study is a mixed, “good news and bad news” picture. All studies that take into account a reasonable sample of field studies show some successes and some failures. About a quarter to a half of the SDGMB studies investigated showed low impact on decision making. On the other hand, roughly half of the studies have led to system-level implemented change with approximately half of the implemented studies being associated with positive measures of success.

“Success” is a multi-dimensional concept in the evaluation studies, having important components on the individual, group, and organization or system level. A promising result is that explanations of differences between successful interventions and less successful ones have not been associated with the complexity of the task environment—of the problems being solved. Rather success appears to be conditioned by team openness to new problem-solving approaches and by an appreciation for and orientation toward empirical problem-solving methods.
Discussion

While recognizing and respecting the difficulties of scientific evaluation of SDGMB, we remain relentlessly optimistic about the method’s utility as a policy design and problem solving tool capable of generating Public Value. Our glass is half (or even three quarters) full. A method that can deliver high decision impact up to three-quarters of the time and implemented results in up to half of the cases examined (and in a compressed time frame) is a dramatic improvement over alternative approaches that can struggle for months or even years without coming to closure on important policy directions.

SDGMB is effective because it joins the minds of public managers and policy makers in an emergent dialogue that relies on formal modeling to integrate data, other empirical insights, and mental models into the policy process. Policy making begins with the pre-existing mental models and policy stories that managers bring with them into the room. Policy consensus and direction emerge from a process that combines social facilitation with technical modeling and analysis. The method blends dialogue with data. It begins with an emergent discussion and ends with an analytic framework that moves from “what is” baseline knowledge to informed “what if” insights about future policy directions. In his dichotomy of model as boundary object versus model as micro-world, Zagonel (2002; 2004) has captured this distinction between SDGMB as a story-telling tool and SDGMB as a analysis tool.

Nearly two decades ago when we began this work, we primarily saw ourselves as technical analysts who were building micro-worlds in an accelerated way, on-line with client groups. Early on we strove to incorporate insights relating to facilitation and group process from our colleagues working in the Decision Techtronics Group (Milter and Rohrbaugh 1985; Quinn, Rohrbaugh,
and McGrath 1985; Reagan-Cirincione et al. 1991; Rohrbaugh 1992; Rohrbaugh 2000). More recently, this work has showed increase sensitivity to stakeholder issues (Bryson 2004b; Bryson, Cunningham, and Lokkesmoe 2002), to issues related to creating visible maps of individual and group thought processes (Bryson et al. 2004), and elements of strategic planning (Bryson 2004a; Eden and Ackermann 1998).

However, we remain convinced that the nature of the models being created in these SDGMB sessions is and will continue to be a key element in the future success of such work. The key to the success of SDGMB interventions is a formal computer simulation model that reflects a negotiated, collaborative, consensual view of the “shared mental models” (Senge 1990) of the managers in the room (the problem structure). This formal simulation model must be tested and tried against existing administrative and time series data (the problem behavior) whenever possible. The final simulation models that emerge from this process are crossbreeds, sharing much in common with data-based social scientific research while at the same time being comparable to the rough-and-ready intuitive analyses emerging from backroom conversations. Furthermore, system diagramming tools both useful for group process and embedded in modern simulation software provide support for dialogues between managers, their mental models, and both structural and behavioral data. In this work, teams join minds; linking people, process, and analysis in the policy design process.

In sum, we believe that a number of the process features related to building these models contribute to their appeal for front line managers:

- **Engagement.** Key managers are in the room as the model is evolving, and their own expertise and insights drive all aspect of the analysis.
• **Mental models.** The model building process uses the language and concepts that managers bring to the room with them, making explicit the assumptions and causal mental models managers use to make their decisions.

• **Complexity.** The resulting nonlinear simulation models lead to insights about how system structure influences system behavior, revealing understandable but initially counterintuitive tendencies like policy resistance or “worse before better” behavior.

• **Alignment.** The modeling process benefits from diverse, sometimes competing points of view as stakeholders have a chance to wrestle with causal assumptions in a group context. Often these discussions realign thinking and are among the most valuable portions of the overall group modeling effort.

• **Refutability.** The resulting formal model yields testable propositions, enabling managers to see how well their implicit theories match available data about overall system performance.

• **Empowerment.** Using the model managers can see how actions under their control can change the future of the system.

Group modeling merges managers’ causal and structural thinking with the available data, drawing upon expert judgment to fill in the gaps concerning possible futures. The resulting simulation models provide powerful tools to ground what-if thinking and contribute to Public Value.

**Notes:**

1. The terms “wicked” (Bryson, Cunningham, and Lokkesmoe 2002) and “messy” (Eden, Jones, and Sims 1983) refer to problems lacking of a clear definition of the problem itself, the goals to pursue or the feasible solutions. Problems are “messy,” partially because of a diversity of stakeholders with different—sometimes conflicting—points of view about a given situation (Bryson and Anderson 2000; Bryson, Cunningham, and Lokkesmoe 2002; Kellog and Mathur 2003).

2. The Decision Techtronics Group has a long tradition developing computer-based models with groups of managers to analyze policy and strategy (Rohrbaugh 1992). DTG’s approach has been used to understand and tackle problems in a diversity of areas using many kinds of models (Mumpower and Stewart 1996; Mumpower, Schuman, and Zumbolo 1988; Milter and Rohrbaugh 1985; Schuman and Rohrbaugh 1991; Larsen and Bloniarz 2000), and particularly using system dynamics models (Reagan-Cirincione et al. 1991; Richardson et al.
1992; Richardson and Andersen 1995; Andersen and Richardson 1997; Kelly 1998; Rohrbaugh 2000; Andersen et al. 2004).

3. GDSS are computer-based systems and techniques developed for group decision support (Quinn, Rohrbaugh, and McGrath 1985; DeSanctis and Gallupe 1987; Nunamaker et al. 1991). The Decision Techtronics Group approach to GDSS, Decision Conferencing, consists of “computer-supported meetings in which several decision makers develop an explicit framework or structure for organizing their thinking about an important, non-routine policy or program choice” (Milter and Rohrbaugh 1985, 183).


5. “Base” policy assumed that all welfare programs were funded at their 1996 levels. The “Base” assumption about unemployment was that the economy was at the exact unemployment rate that would cause no growth, but also no decline over the time horizon 1984-1998 (the modeling team “backwards computed” this figure as part of the model-testing and confidence-building phases. This calculation was intended to “hold constant” the very large effects of unemployment on TANF caseloads so that the runs could show “pure” effects of policy changes.

6. Note that in order to show the pattern of the dynamics, the scales on figures 1 and 2 are not zero-based scales and the scales are different. Thus the visual intervals between graphs in figures 1 and 2 are not comparable.

References


