Towards evaluation of systems thinking interventions: a case study

Steven Cavaleri* and John D. Stermanb

Abstract

The use of system dynamics and systems thinking in organizations is increasing dramatically. However, little is known about the efficacy of these interventions in enhancing organizational effectiveness or productivity. With few exceptions, the relationship between the use of systems thinking and organizational performance remains the province of anecdote rather than rigorous follow-up research. Evaluation helps identify the tools and techniques that can be used successfully, the situations in which they work, and their limitations. This article reports a follow-up evaluation of a well-known systems thinking intervention designed to improve quality and performance in the claims adjusting unit of a U.S. insurance firm. Through a questionnaire, interviews, and extensive review of company records, it has been established that the intervention succeeded in changing the mental models and behavior of key managers. New policies consistent with model recommendations were implemented. Evidence of performance improvement is mixed, however. A number of confounding variables limit the possibility of assessing the business impact of the intervention. These limitations could have been lessened if the original intervention had been designed with rigorous evaluation as a goal. © 1997 by John Wiley & Sons, Ltd. Syst. Dyn. Rev. 13, 171–186, 1997

Introduction

Recent innovations in systems thinking have fueled growing interest among managers in the practical application of the tools (Senge 1990; Morecroft and Sterman 1994; Senge et al. 1994). A number of technologies and protocols are useful for developing systems thinking capability in organizations and individuals, such as management flight simulators, experiential exercises, and causal loop diagramming. Although various intervention techniques that fall under the rubric ‘systems thinking’ have become quite popular, little is known about their efficacy in enhancing organizational effectiveness or productivity. With few exceptions the relationship between the use of systems thinking and organizational performance remains the province of anecdote rather than rigorous follow-up research. We argue that such rigorous follow-up research is essential to build a strong foundation for the refinement and wise use of the tools of system dynamics and systems thinking—a goal of both academics and practitioners. The need for rigorous evaluative research is widely accepted; for example, Strake (1967) writes of...
educational innovations: “Folklore is not a sufficient repository. In our data banks we should document the causes and effects, the congruence of intent and accomplishment, and the panorama of judgments of those concerned” (p. 539). At the same time, evaluative research is complex and poses many challenges. In this article we report a follow-up study on a well known and often cited systems thinking intervention in an organization. The evaluation suggests that the intervention did benefit the organization and its employees. However, because the original intervention was not designed with evaluation in mind, we were forced to conduct a retrospective study, greatly complicating the task of assessment and limiting our ability to learn from an innovative use of system dynamics to aid senior management. We return to this theme in the discussion.

The history of the intervention is well told elsewhere, primarily in Senge (1990). Bergin and Prusko (1990) provide an account from the perspective of two key company participants; see also Senge and Sterman (1992) and Moissis (1989). We summarize briefly here. In the mid-1980s an experienced modeling team assembled by the MIT System Dynamics Group worked with a top management team of a mid-size property-casualty insurer to develop a system dynamics model describing the interaction of claims management, quality, and costs. The model suggested important insights into the industry-wide problem of rising costs and falling quality, including some high-leverage policies to improve the situation. To diffuse the insights from the model more widely throughout the organization, the model was converted into an interactive management flight simulator, the “Claims Game”. The flight simulator was incorporated into a “learning laboratory”, in which the participants were introduced to various systems thinking tools, play of the “Beer Game” (Sterman 1989), a workshop with the claims game, and a seminar which dealt with systems thinking skills, such as causal-loop diagramming. The learning laboratories were initially run at corporate headquarters with participants from different regional offices and functions, but were eventually devolved to individual regional offices, where intact management teams could participate. The program began in 1988 and continued through 1991.

Data sources and level of analysis

We sought to evaluate the impact and effectiveness of the intervention. We focus on assessing changes in attitudes, practices, and business results subsequent to the training program. Our research takes a “formative perspective” (Gagne 1985), in the sense that generally accepted tools and processes for evaluating the impact of systems thinking on organizational performance do not yet exist. To enhance reliability and ensure we captured as many perspectives as possible, we used
multiple data sources, including a questionnaire, interviews, and extensive review of primary archival data such as personnel records and operating results.

Our study was conducted in a regional office of the company. The primary function of the office, which employs about 70 people, is claims management. All members of the regional office participated in at least some of the training. In our evaluation, we considered all employees of the regional office, but distinguish between managerial and non-managerial personnel because the managers received more extensive training than the non-managers. All employees attended the Beer Game module, but the Claims Game flight-simulator session and systems thinking seminar were attended only by members of the claims unit along with the office's top management team.

We began by interviewing the key managers of the regional office. The interviews were used to convey the goals of the research and to identify indicators for assessing changes in performance. We also solicited the managers' views regarding the usefulness of the training program. A 22-item questionnaire was designed and distributed to all of the participants in the training who were still with the organization at the time of our study (full documentation is available from the first author). The questionnaire was designed to measure changes in the systems thinking capabilities of the participants, including any differences in personal perceptions or behavior since the training, and also to measure their ability to recognize specific systems principles demonstrated in the Beer Game and systems thinking training sessions. Following each question opportunities were provided for respondents to offer clarifying comments.

Hypotheses

Three claims have been made in the literature about the effectiveness of systems thinking interventions: that they alter thinking, behavior, and results (e.g. Senge 1990; Senge et al. 1994). In the context of this work, the first assertion is that the insurance claims management learning laboratory should have altered people's mental models to be more systemic and to clearly capture the long-term dynamics of the business. The hypotheses below distinguish between the managers and the other workers in the regional office we examined. Thus:

H1-A. Managers attending the training will have an increased capability to think systemically.
H1-B. Non-managerial participants will have an increased capability to think systemically.
Second, changes in behavior consistent with the long-term high-leverage policies identified in the simulation analysis should also be observed. Thus:

H2-A. Managers attending the training will change their behavior and organizational policies to be more consistent with the long-term best interests of the system as a whole.

H2-B. Non-managers attending the training will change their behavior to be more consistent with the long-term best interests of the system as a whole.

Finally, it is expected that changes in behavior will improve business results. Thus,

H3. The claims unit’s operational performance will improve as a consequence of the training.

We discuss the metrics of operational performance and their suitability below.

Results

Participant attitudes and beliefs (survey results)

The questionnaire was given to all employees of the office who participated in the systems thinking intervention (59 out of a total of 70 employees). Thirty-six usable responses were received (61%); nine were from managers and 27 were from non-managers. The response rate is excellent; the main reason for non-response was that some personnel were on vacation at the time the survey was administered; others no longer worked at the firm and were not accessible. The survey results are organized in four areas:

1. self-reported cognitive changes;
2. self-reported behavioral changes;
3. perceived changes in the way the company is managed;
4. competency in understanding the principles demonstrated in the Beer Game (Figure 1).

Self-reports of cognitive change. The effect of the training on self-reports of systemic thinking differs between managers and non-managers. Non-managers, who received only the Beer Game component, report little awareness of changes in
Fig. 1. Questionnaire results
their thinking. Managers, who took the entire training, report much greater awareness of changes in their thinking, which they attributed to the intervention.

**Self-reports of behavioral change.** Again, there is a strong difference between managers and non-managers. Non-managers report no change in behavior, while managers believe their behavior changed as a result of the intervention. The interviews strongly suggest that managers believe their behavior change was a direct effect of the training.

**Perceived changes in management style.** The survey data provide evidence for differences in perceptions of the managerial style of senior office personnel. Non-managers noted only a slight change in their perception of management's style and the degree to which it became more systemic. However, the managers report that their management style became more "systemic".

**Understanding the key principles of the Beer Game.** In the questionnaire, respondents were asked to read a series of short case descriptions and rate the extent to which the lessons of the Beer Game were illustrated or relevant to the case. Responses were compared against the average rating of a panel of three experts. The scores ranged from 0.09 to 0.58 (where 0 would indicate perfect concordance with the experts), and 0.67 is the rating achieved if responses were random. The overall mean was 0.28. The mean for the managers was 0.19, while for non-managers it was 0.31. Participants appear to have learned some, but not all, the lessons of the Beer Game. Managers did better than non-managers.

**Manager comments**

The interviews conducted with the management team and the written comments from the questionnaire were quite revealing. Respondents report that they developed new insight into the causal relations at work in their unit by playing the Claims Game. In particular, the visualization of the interaction between the stocks and flows within the claims unit helped in the design of new policies. Several managers noted that playing the claims game helped to reinforce their prior understandings of systems thinking. It also helped them to understand the interconnectedness of elements in the system and to see the tradeoffs generated by alternative courses of action. Managers reported that the systems thinking training program helped shift many people's thinking from a reactive to a more strategic mode, and believed this gave them an edge over competitors who relied on a traditional view of managing.

In sum, managers reported that the training affected their mental models (attitudes and beliefs) and that they understood and were able to apply the
principles of systems thinking to their jobs and in their interactions with colleagues. Other personnel did not report such changes.

Unfortunately, it is not possible to say whether the differences in the responses of managers and non-managers are due to differences in the amount of training in systems thinking they received or to other, intrinsic differences between the managers and other personnel. The design of the original intervention, where all managers received one type of training and all others received another fully, confounds the manager/worker distinction with the full-training/partial-training distinction. Note that the confound could have been avoided if each group had been divided into treatment and control subgroups.

Furthermore, in interpreting these data one must be sensitive to the possibility of demand effects, that is, of being told what the informants believe you want to hear. It is possible that the managers believed we were looking for a positive outcome, or that they had an investment in a positive outcome since they had committed substantial time to the training process. Such demand effects are a common problem in field work, especially when relying on retrospective reports.

Changes in actual behavior

We next sought to document specific changes in behavior flowing from the intervention. These artifacts of change would show the tangible impact of the learning laboratory directly. Our interviews and archival data revealed a range of changes in behavior, policy, and organizational structure (Table 1). Many of the changes can be traced directly to the training. In interviews, managers were explicit in relating these new policies to their new mental models of the causal structure of the claims unit. They developed several new strategies to implement key recommendations emerging from the model, including recruiting more experienced and higher quality adjusters, attempting to retain experienced adjusters longer, focusing on the quality of the settlements rather than measures of throughput, and increasing total settlement capability by hiring new adjusters (Senge and Sterman 1992).

As examples of these changes, consider the hiring and staffing policies. The simulation model suggested that long-term, system-wide costs could be reduced by increasing the organization’s capacity to settle claims and then using the extra capacity to increase the quality of settlements rather than increase the throughput. To increase capacity requires increasing the skill and experience of adjusters, as well as the total headcount. Headcount limits for the regional offices are set by corporate headquarters. The regional office did request an increase in their authorized headcount cap, but the request was denied. The office managers cleverly realized that they could increase effective adjusting capacity without an increase in
Table 1. Summary of behavioral changes and new policies

I. Company policies
   1. New statement of performance expectations developed.
   2. Work quality redefined.
      Quality was emphasized more compared to throughput, and quality was redefined to include the long-term and system-wide effects of claimant contact, negotiation, documentation, and investigation.

II. Structure
   1. Supervisor converted into adjuster.
   2. Adjuster jobs redefined.
   3. Cases assigned to adjusters in new way.
      Claims were assigned to adjusters to match expertise better to the complexity of the claims.
      The result was faster case resolution with less stress.

III. Hiring
   1. New selection criteria implemented, stressing those with an aptitude for systemic thinking.
   2. New recruitment methods implemented, in particular pre-recruitment networking with experienced adjusters throughout the industry was instituted.
   3. Time to fill vacancies reduced.
   4. Ideal candidate profile redefined.
   5. Interviewing process changed (new questions added to select candidates with systemic thinking capability).
   6. Hiring experienced adjusters emphasized.
   7. Addition to staff of adjusters requested.
      These measures resulted in a substantial increase in net adjuster headcount and effectiveness, even though the request to corporate headquarters to increase the authorized headcount (#7) was denied.

IV. Training
   1. Goal of training altered to stress adjuster empowerment; adjusters were given the ability to spend time to maintain quality in face of pressure to meet throughput goals.

V. Environment
   1. Outreach to attorneys, explaining systemic issues, to help legal staff understand the changes in policies at the claims office.
   2. Systematic follow up with customers implemented to get more rapid feedback on service (satisfaction survey cards mailed to all claimants after settlement).

the authorized headcount of the office by filling the vacancies created by turnover faster. Before the learning laboratory, management's goal was to minimize administrative expenses by delaying the replacement of departing employees. The average number of adjusters was well below the authorized headcount. After the intervention, the goal became rapid replacement of departed employees, effectively increasing headcount and adjuster capacity without requiring corporate approval.

To test this claim, we calculated the average time to fill vacancies pre- and post-intervention directly from the office's personnel records (Table 2). The average time to fill vacancies declined by nearly a factor of four after the intervention, strongly consistent with the managers' reported change in policy.

The organization also sought to reduce worker burnout and the consequent
increase in turnover, thus stanching the outflow of expertise. The decision rule for case assignments was changed from a simple queuing system, in which claims were assigned to the next available adjuster, to one in which the complexity of claims was better aligned with adjuster experience. This policy was designed to prevent stress and work overload for inexperienced adjusters and give experienced adjusters more challenging and satisfying work. It took courage to implement this policy. Despite the enormous long-run benefits of retaining experienced people, with their greater productivity and skill, in the short run the new policy meant that a claim of given complexity might have to wait until an adjuster with suitable experience became available, even if others without the right mix of skills were free, eroding customer service and raising administrative cost per claim. This worse-before-better dynamic is typical of the high-leverage policies identified by the modeling work and a primary reason that these policies had not already been implemented (Senge and Sterman 1992). Turnover did decline after the intervention, falling from 27%/year in 1988 to 9%/year in 1991 (turnover rose substantially in 1992, however, see Table 3).

Two notes of caution: First, some of the turnover is desired, as it is necessary to weed out poor performers. We have no way to distinguish between “desired” and “undesired” turnover. Second, the decline in turnover, while consistent with the implementation of the new policies, could also be explained by exogenous events: the decline in turnover coincided with the national recession in 1990–1991. During recessions, voluntary quits drop as workers find it harder to land new jobs. Thus the decline in turnover could have been the result of changing macroeconomic conditions. It is simply not possible to rule this out with the small sample size and limited data available.

Changes in business performance

In the end, the effectiveness of any intervention rests on changes in business outcomes (the so-called “system improvement test”; see Forrester and Senge 1980).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TURNOVER (%/YEAR)</td>
<td>27%</td>
<td>18%</td>
<td>18%</td>
<td>9%</td>
<td>27%*</td>
</tr>
</tbody>
</table>

*9 month period, January–September 1992

Table 2. Average time to fill vacancies in claims adjuster positions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AVERAGE TIME TO FILL VACANCIES (WEEKS)</td>
<td>16.6 weeks</td>
<td>4.6 weeks</td>
</tr>
<tr>
<td>RANGE</td>
<td>5–20 weeks</td>
<td>1–12 weeks</td>
</tr>
</tbody>
</table>

Table 3. Rate of employee turnover
To be judged effective, an intervention must do more than affect attitudes and beliefs; it is not enough that participants enjoy the training and rate the workshops highly. An intervention must also have positive effects on the states of the system. The performance of the claims unit was measured by four standard measures used throughout the industry and routinely tracked within the company (Table 4). These are:

1. the claims pending ratio (the ratio of pending [unsettled] claims to the flow of new claims);
2. the production ratio (the ratio of claims settled to incoming claims);
3. average settlement size ($ per feature; a feature is a particular loss described in a claim [complex claims consist of many features]);
4. average administrative cost per feature.

The claims unit’s performance was examined over a six-year period from 1986 through 1992. During this time the number of incoming cases grew only slightly, by 4%. However, there was a qualitative increase in the complexity of the cases, as judged by the managers. In interpreting the data, note that the rate of incoming claims is subject to considerable year-to-year variation. There is simply too little data available to conduct statistical tests.

The pending ratio measures the average time required to settle claims. There is no definitive change discernible in this performance measure. The pending ratio improved 13% from 1989 to 1991. However, in 1992 it rose to 2.65 from 2.38 months the previous year. Managers reported that the target pending ratio is between 2.00 and 2.25 months, depending on the mix of business. It becomes increasingly difficult to make further gains as the ideal ratio is approached.

<table>
<thead>
<tr>
<th>Year</th>
<th>Pending ratio</th>
<th>Production ratio</th>
<th>Settlement size ($)</th>
<th>Administrative cost per feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>2.67</td>
<td>102.5</td>
<td>1432</td>
<td>115.25</td>
</tr>
<tr>
<td>1987</td>
<td>2.45</td>
<td>98.4</td>
<td>1639</td>
<td>143.88</td>
</tr>
<tr>
<td>1988</td>
<td>2.75</td>
<td>98.0</td>
<td>1895</td>
<td>166.93</td>
</tr>
<tr>
<td>1989</td>
<td>2.43</td>
<td>98.3</td>
<td>1597</td>
<td>168.19</td>
</tr>
<tr>
<td>1990</td>
<td>2.69</td>
<td>100.1</td>
<td>2102</td>
<td>195.54</td>
</tr>
<tr>
<td>1991</td>
<td>2.41</td>
<td>98.8</td>
<td>1711</td>
<td>198.49</td>
</tr>
<tr>
<td>1992</td>
<td>2.46</td>
<td>99.92</td>
<td>2333</td>
<td>246.42</td>
</tr>
<tr>
<td>1993</td>
<td>2.47</td>
<td>100.84</td>
<td>2529</td>
<td>279.03</td>
</tr>
<tr>
<td>1994</td>
<td>2.65</td>
<td>97.68</td>
<td>3222</td>
<td>274.24</td>
</tr>
</tbody>
</table>
The production ratio is the ratio of claims settled to claims received, and should not be less than 100% (in equilibrium incoming claims = settlement rate). No clear pattern of improvement is evident. Over a six-year period performance was both above and below the overall company average. By 1992 the ratio had fallen 1.4% below the prior year and 5.6% from the level of 1990. Comparisons to national averages are problematic. The incidence of losses that drive the incoming claim rate and thus the performance metrics are not correlated perfectly throughout the regions of the country in which the company writes business.

Average settlement size for a given time period is calculated by dividing the total payout for claims settlement by the total number of features settled during the same time period. Again, the data suggest no definite pattern of improvement. Payouts to customers in this regional office have typically been higher than the national average because of the higher cost of living and of auto repair in the states served by the office. Payouts made by the office declined in 1991, but increased substantially in 1992. Settlement size is influenced by regional factors, such as cost of living, weather, and economic conditions. Substantial time delays may arise in reporting the settlement of cases as a result of accounting conventions (a case is not credited as settled until all features are resolved). The rise in costs in 1992 may also be explained by the change in the mix of claims for this office or by changes in underwriting standards. These exogenous variables further confound the interpretation of the results.

Settlements involve a direct payout of funds to claimants, but also incur administrative costs such as appraisal fees, adjuster salaries, legal fees, and overhead. Attorney’s fees are the largest component of this index. It was believed that increasing the quality of claims adjusting would reduce the need for legal intervention, i.e. litigation and subrogation, reducing expenses (Senge 1990). However, administrative costs show no improvement; the office’s performance runs roughly parallel to the national company average; both increased steadily during the period considered.

Externalities: confounding variables

Interpreting the performance data is difficult because, as with all longitudinal studies of organizational change, the environment inevitably changes along with the changes introduced by an intervention. Such is the case here. In particular, a number of substantial changes in the environment of the firm, the insurance industry and the firm itself confound the interpretation of the data above.

Changing mix of business. The types of claims handled by the office have changed in two significant ways since the intervention. One of the shifts has been
evolutionary, the other is more sudden. The claims handled by the office became more complex. In general, they included more features and were more frequently the subject of litigation. Both of these changes reflect long-standing trends in all areas of the U.S. insurance industry. Second, the branch office was given responsibility for processing claims on policies issued by a recently acquired business unit. The new unit tended to write policies in environments involving greater risk and requiring more time to settle. A higher portion of its claims involve personal injury claims, which tend to be heavily litigated. No data were available to assess the magnitude of this effect.

Change in Management. Subsequent to the start of this research, the parent corporation experienced several large scale changes in ownership and management. Controlling interest in the firm was acquired by a former minority stockholder and there was significant turnover in the top management team. The policies advocated by the new top management team were perceived by some members of the regional office as running counter to those emerging from the systems thinking intervention. The new top management team of the corporation emphasized cost reduction as a primary focus, rather than quality improvement, as the systems thinking intervention stressed. The impact of this shift in corporate strategy on the regional office is difficult to assess. A number of the new policies established following the training were “reassessed” in light of the new management position; several key managers subsequently received attractive offers from other insurance firms and left the company.

Discussion

The hypotheses offered above suggest that the claims management learning laboratory should have altered the way people think about their business so that it became more systemic, altered their behavior to enact high-leverage policies for system-wide, long-term improvement, and, as a result, measurably improved business results.

Our evaluation of the systems thinking intervention yields mixed results. The questionnaires, written comments, and interviews do support hypothesis 1-A, that managers did experience a shift in their mental models towards a more systemic understanding of the claims systems and its dynamics. The managers who participated in all of the training modules believe that their thinking shifted significantly towards a more systemic style. However, there is no evidence to support the idea that the intervention helped non-managers, who received much less training, to think more systemically (H1-B is not supported).

The evidence for behavioral change is again mixed. There is little evidence of
behavior changes among the non-managers (H2-B is not supported). It is, of course, not surprising that simply playing the Beer Game and attending an introductory lecture on systems thinking had little effect—changing the deeply held mental models that drive behavior is difficult. In striking contrast, there is no doubt that behavior did change among managers (H2-A is supported). The management team of the regional office designed and implemented new policies for recruiting, hiring, training, and retaining adjusters. Changes were made in the nature of the work, in the assignment of work to adjusters, and in the emphasis on quality compared to throughput. The policy changes implemented by the business unit managers were creative, synergistic, and clearly oriented to long-term improvement. For example, policies designed to reduce turnover and shorten the time to fill vacancies reinforced one another. As a result of cutting turnover from three adjusters per year in 1988 to one per year in 1991, with a simultaneous reduction in the time to replace these departures from about 17 weeks to about 5 weeks, net available adjuster capacity increased by nearly a full time equivalent, a substantial boost in an office with about 11 adjusters. More important, the managers’ new policies and behaviors took courage: the managers consciously accepted a short-run deterioration in important performance metrics that corporate headquarters used to evaluate their performance.

Despite the significant changes in policies and behavior, there is minimal support for the claim that the intervention produced measurable improvements in business performance during the six years following the intervention (H3). None of the pending ratio, production ratio, settlement costs or expense ratios shows any consistent patterns of improvement. Average adjuster productivity (features settled per year per adjuster) did rise from 1986 to 1991 by 14%, but fell by 11% in 1992, the same year turnover increased. This pattern of improvement is suggestive of benefits that may be related to the intervention, but also calls into question management’s ability to limit turnover and sustain the improvement.

The weak evidence of performance improvement seems at odds with the strong evidence of change in manager’s mental models, behavior, and organizational structure and policy. However, the absence of evidence is not necessarily evidence of absence; advocates and critics of the intervention can both interpret the results to be consistent with their positions. Since the intervention was not designed as a controlled experiment, an advocate could seek to save hypothesis H3 in the face of the negative data by arguing that performance would have been even worse without the training, or that it is too early to see the beneficial effects. Second, the model suggested a short-term/long-term tradeoff in which performance suffers between the time the organization invests in the quality of its adjusters and the time the effects of superior quality manifest. As illustrated above, policies designed to reduce turnover, enhance adjuster experience and thus boost productivity and
processing capacity actually reduce throughput in the short run, lowering the pending ratio and raising the administrative cost per feature. Third, the benefits of some policy changes may not be captured by the standard metrics, such as cost savings in the human resources department due to reduced turnover, or increased customer loyalty resulting from the higher quality work of more experienced and motivated adjusters. Alternatively, a critic can argue that the lack of performance improvement is because the theory embodied in the original simulation model is wrong, so that the changes in attitudes, behavior, and policies had no or even a negative effect on performance.

Despite the relatively long time that has transpired, the use of multiple data sources, and the extensive cooperation we received from the personnel at the regional office, it is simply not possible to rule out these competing hypotheses. We point out that the differing possible explanations for the apparent lack of performance improvement to illustrate the complexity and ambiguity inherent in an effort to substantiate claims of efficacy offered by those who conduct an intervention. The ambiguity is compounded when an intervention is not designed as an experiment or as a prospective study. Nevertheless, the burden of proof must rest on those who would argue for the efficacy for the intervention. We note that the impact of the intervention is observable primarily in the domain of policy and perception, raising questions about the relationship of these factors to performance.

However, while evidence for the impact of the intervention on performance is mixed, the dynamic hypothesis developed in this study has received strong support. Oliva (1996) tested the theory in a similar service setting (retail lending in a large UK bank). Oliva was able to overcome many of the data limitations here and successfully estimated statistically the key behavioral relationships in the model. Oliva's study provides strong support for the dynamic hypotheses developed in this work about the sources of poor service quality and rising costs in financial services and other high-contact service settings.

Methodological Conclusions

Ambiguity in the assessment of the changes in business performance highlights the main conclusion we draw from this study. Impediments to learning about complex systems — such as a systems thinking intervention in a large organization — are well documented (Sterman 1994). The difficulty of attributing business results to particular interventions in a complex dynamic system is both a thorny problem in evaluative research and a chief cause of superstitious learning and the perpetuation of dysfunctional organizational polices. The lack of baseline data taken in real time
before and during the intervention intensified these difficulties, limiting our ability
to discriminate among competing hypotheses about its impact.

It is understandable that proper design and resources for evaluation and
assessment suffer when action researchers begin a new project. Identifying
potential partner organizations, negotiating entry, building trust, and working with
the partner team to understand the business dynamics are difficult and demanding
tasks. Often, members of organizations seeking help do not appreciate the benefits
of research and are reluctant to provide the resources to support it, or to permit
basic protocols such as treatment and control groups or randomized assignment.
Our study shows how such behavior leads to a short-term/long-term tradeoff. The
experience at this firm did generate important benefits, including the Microworlds
software (Diehl 1994), experience with the design and delivery of “learning
laboratories” and a rich theory of service quality dynamics (Senge and Sterman
1992; Oliva 1996). But, while we uncovered tantalizing suggestions that the
systems thinking intervention had a significant impact on the organization, the
evidence is weakened because the evaluation was retrospective. As Laird (1985)
notes, “evaluation begins in the design phase—not after the program has ended”
(p. 267). It is encouraging that an increasing number of system dynamics
interventions have been designed to facilitate evaluation from the start (see, e.g.,
Huz et al. 1997). Despite its difficulties, the artful marriage of action-oriented
interventions and rigorous follow-up study is vital if the potential to learn about the
dynamics of systems thinking in organizations is to be fully realized.

Note

1. The team consisted of Ernst Diehl, Nathan Forrester, Daniel Kim, and Peter
Senge.

References


Learning Organizations, ed. J. Morecroft and J. Sterman, Portland, OR: Productivity


Winston.


