Empirical Validation of a Dynamic Hypothesis

Rogelio Oliva

Sloan School of Management, MIT

30 Memorial Dr., Room E60-355 Cambridge, MA 02142

Tel 617/253-0834   Fax 617/252-1998   e-mail: roliva@mit.edu

The purpose of this paper is to describe the methodological approach followed to validate a dynamic hypothesis of service delivery and explain its implications for service quality. For a full report on the application of the methodology and the substantial results obtained in the analysis see Oliva (1996).

Background. The starting point for this research is a dynamic hypothesis – a potential explanation of how structure is causing observed behavior – of the interactions between service capacity and service quality that was articulated in the context of a multiple-year system dynamics study with Hanover Insurance Company (Senge, 1990; Senge and Sterman, 1992). In the six years since the original theory of service delivery was developed in the insurance context, the model has been recast as a generic theory for high-contact services (Oliva, 1993b; Senge and Oliva, 1993), turned into a flight simulator (MicroWorlds, 1994; Oliva, 1993a) and used in workshops for hundreds of managers from diverse service industries. From this experience, it was speculated that the findings from the Hanover Insurance case are applicable to a wider set of service settings.

Unfortunately, most of the research work done in the quality of goods arena has proven inadequate for understanding service quality. Fundamental differences in the way services are produced, consumed and evaluated make the lessons from the literature on quality and consumer behavior inoperative in a service context (Zeithaml, Parasuraman and Berry, 1990). Researchers from the operations management, human resources and marketing have dedicated considerable efforts to explore the main determinants of service quality. Although some integrated frameworks of service delivery and service quality have been articulated, most of the evidence available for the relationships proposed in these frameworks is fragmented. The purpose of this research was to develop and test an integrated theory of service delivery capable of generating insights into the challenges of managing service quality.

Approach. The research activities can be grouped into three distinct stages:

1. Formalization and Substantiation of Theory. The proposed theory of service delivery integrates findings from different disciplines that have examined the service delivery process. The theory, while being grounded in the human resources, behavioral decision theory, marketing, and operations management literature, was articulated using a system dynamics model along with a detailed account and evidence from the literature for the proposed constructs, causal linkages, and formulations that compose the theory. A computer simulation model can be an effective tool for validating theory. First, the model formalizes the hypothesized relationships between variables creating a refutable causal model with multiple 'points of testing' (Bell and Senge, 1980). Second, it enables testing of the completeness and coherence of the proposed relationships.

2. Empirical Validation of the Theory. Although the proposed theory describes the relationships between variables throughout the service setting, much of the evidence available for those relationships is fragmented and specific to the relationships. In testing a complex dynamic theory, there are three validity concerns that should be addressed:

- Does the micro-structure of the model correspond to what is known about the real system?
- Do the estimated or observed relationships support the theory?
- Can the macro-behavior of the service setting be explained from the structural components of the theory?

These concerns guided a validation strategy based on calibrating the existing model of service delivery to fit the structure and behavior of a service setting. Calibration of a model to an empirical setting attests to the model's capability of capturing the characteristics of the of the

research site and its potential relevance to managers. Although it is impossible to verify a model (Oreskes, Shrader-Frechette and Belitz, 1994), insofar as the proposed formulations are capable of capturing the behavior observed in a service setting we can augment our confidence in the theory. The selected service setting was a back-office center in a major British bank responsible for making loan decisions for the mass market and small business accounts.

To address the structural validity issue the calibration was done through partial model estimation with immediate data sources. The process involved a combination of detailed field study, analysis of numerical data, and formal model development. Other validity concerns were addressed through a suite of tests performed at the full system level. A brief description of the calibration strategy and the full system tests is given in the following sections.

3. Derivation of Managerial Implications. The findings from the validation process were used to generate insight into the relative strength of the different responses to work pressure and to propose a more parsimonious and empirically appealing formulation for the formation of service aspirations. A second set of managerial implications was the identification of leverage points and policy recommendations for managing quality in a high-contact service setting. Finally, to facilitate the generalization and transferability of insights, the model was taken outside the high-contact service context and its usefulness in other service settings explored. By explicitly examining the application domain of the theory – the set of structures and behaviors the theory is capable of explaining – it was possible to define a generic framework to link structural characteristics of service settings to the problematic dynamics observed in the service industry.

Calibration Strategy. Forrester's distinction between overt and implicit decisions (1961) was used to develop a calibration strategy. Calibration of *implicit* decisions, or the parameters that drive them, was limited to identifying – through observation or interviews – the physical attributes of the workflow in the research site. Alternatively, the majority of the calibration efforts were focused on the statistical estimation of the parameters describing the model's *overt* decisions and the information processing capabilities of the agents in the service setting (Graham, 1980; Mass and Senge, 1980; Senge, 1977).

For each decision or set of parameters of interest, 'detailed data,' i.e., data specific to the relationship under study, were collected from the field site and the parameters or shape of the relationships estimated through non-linear least squares estimation using Powell's (1969) optimization algorithm as implemented in Vensim® (Ventana Systems, 1995). Analysis of the residuals from the estimations was extremely helpful in discovering subtle flaws in the initial formulations. In case of a lack of field data to test a micro relationship, I adhered to the system dynamics paradigm and incorporated in the model the best estimate available from the existing literature and previously available empirical research (Forrester, 1975).

Full System Tests. Replicative validity was tested through the model's ability to match the historical behavior of the lending center. The dynamic significance of the structural components was tested through sensitivity analysis. Finally, extended simulations were used to test the overall dynamic hypothesis articulated by the theory.

Historical Fit of the Model. To test the historical fit of the proposed theory, the model was simulated with two exogenous data series driving it: the weekly demand on the lending center and the weekly rate of absenteeism. Both of these series had a significant random component and were outside the model boundary. The summary statistics for the historical fit of the model to six data series – desired labor, total labor, time available to process orders, orders processed, time allocated per order and work intensity – were calculated (data was weekly and was available for one year). The Mean Absolute Percent Error between the simulated and actual variables was less than 2% for all series indicating a close fit of the model to the actual behavior of the lending center. Low bias and variation components of the Theil inequality statistics indicated that the errors were unsystematic (Sterman, 1984).

Significance of Behavioral Components. To test whether the observed system behavior was being generated by the hypothesized causes, the overall dynamic hypothesis was broken down

into four behavioral components – management hiring policies, employees' learning curve, employees' response to work pressure and effects of perceived quality on performance. Sensitivity analyses were performed through a set of simulations varying system parameters that affect the strength each of these elements. Despite the confounding effects of the transient behavior remaining from the buildup stage of the lending center, enough evidence was found to corroborate each of the behavioral components of the proposed theory and their impact on the center's performance.

Extended Simulations. To assess the implications of the current policies of the lending center under stable conditions, the simulation horizon of the model was extended for two years beyond the final point where data were available. Since the two data series driving the model did not show any significant trend component, it was possible to capture their main characteristics with a pink noise random number generator capable of reflecting the same variance and autocorrelation spectrum (Britting, 1973). The extended simulations showed that, as predicted by the theory, the structural elements of the research site – policies and physical flows – bias its performance towards an erosion of service quality.

Generalizing the Theory. To assess the transferability of insights and recommendations derived for the high-contact service sector it was necessary to address the issue of external validity of the theory – "the extent to which one can generalize the results of a research to the populations and settings of interest" (Judd, Smith and Kidder, 1991, pg. 28). External validity was explored in two dimensions: the range of behaviors and reference modes that the theory is capable of explaining, and the variety of service settings that can accurately be captured by the proposed structure. The two dimensions – behavior and structure – define the application domain of the theory (model).

The variety of reference modes that can be generated by the model was explored by varying system parameters. The characteristics of service settings that can be captured in the model were identified and grouped into the factors affecting the potential responses that a service setting could have to environmental changes. The identification of the main characteristics of service delivery process not only allowed exploration of the flexibility of the model to capture other service settings, but it also permitted the identification of the characteristics that define the space where particular policy recommendations are valid. Finally, the model structure and the generalized response mechanisms were used to link structural parameters of service settings to the problematic dynamics observed in the service industry.

Discussion. One of the long standing claims of system dynamics has been that of generalizability, i.e., the creation of a common frame of reference to capture the characteristics of a system and make them transferable to other settings (Forrester, 1961). The kernels of transferable knowledge in the system dynamics field have been captured as 'generic structures' and Forrester's claim "... that about 20 such general, transferable ... cases would cover perhaps 90 percent of the situations that managers ordinarily encounter" (1993, pg. 210) testifies to their perceived importance in the development of the field.

Model validation has been one point on which system dynamicists and other disciplines disagree. In the SD tradition, validation has focused on construct and internal validity, but has not explored the dimension of external validity. Although construct validity and internal validity are prerequisites to external validity, without addressing the issues of external validity it is impossible to make the generalizability claim, and, therefore, it is quite difficult for 'generic structures' to become part of mainstream management theory. The approach followed in this work is presented as the first steps for a methodological strategy to address the validity issues of system dynamics models.

Formalization of behavioral models as in a system dynamics model normally constitutes an excellent proof con *construct* validity. Matching historical behavior only tests the *replicative* validity of a model. A full test of model's representativeness has also to consider its *structural* validity (face validity). The derivation of the model structure and parameters from observed

micro-decisions and physical flows in the service setting – data obtained through interviews and field studies – and the ability for partial model structure to replicate intermediate data series constitute true tests of the model's structural validity.

The issues that need to be addressed when exploring *internal* validity are if the observed behavior is indeed caused by the structure that has been specified, and if the structure, as calibrated by the partial-model estimation process, is capable of generating the hypothesized reference mode. The sensitivity analysis and the extended simulations were used to explore these issues. Finally, *external* validity of the theory was ascertained through a rigorous exploration of the application domain of the theory.

Although the validation strategy was developed with the idea of testing a preexisting theory in a real world situation, the same strategy could be used to test dynamic hypotheses in a traditional system dynamics intervention.

References

Bell, J.A. and P.M. Senge. 1980. Methods for Enhancing Refutability in System Dynamics Modeling. *TIMS Studies in the Management Sciences*, **14** (1), 61-73.

Britting, K.R. 1973. *Correlated Noise Generation Using DYNAMO*. System Dynamics Group, MIT. D-1908.

Forrester, J.W. 1961. *Industrial Dynamics*. Cambridge, MA: MIT Press.

Forrester, J.W. 1975. The Impact of Feedback Control Concepts on the Management Sciences. In *Collected Papers of Jay W. Forrester*. (pp. 45-60). Cambridge, MA: Productivity Press.

Forrester, J.W. 1993. System Dynamics and the Lessons of 35 Years. In K.B. De Greene (Ed.), *Systems-Bases Approach to Policymaking*. (pp. 199-240). Norwell, MA: Kluwer Academic Publishers.

Graham, A.K. 1980. Parameter Estimation in System Dynamics Modeling. In J. Randers (Ed.), *Elements of the System Dynamic Method*. (pp. 143-161). Cambridge, MA: Productivity Press.

Judd, C.M., E.R. Smith and L.H. Kidder. 1991. *Research Methods in Social Relations*. Fort Worth, TX: Holt, Rinehart and Winston, Inc.

Mass, N.J. and P.M. Senge. 1980. Alternative Test for Selecting Model Variables. In J. Randers (Ed.), *Elements of the System Dynamic Method*. (pp. 203-223). Cambridge, MA: Productivity Press.

MicroWorlds. 1994. *Service Quality Microworld*. Cambridge, MA: MicroWorlds, Inc.

Oliva, R. 1993a. *Service Quality Management Flight Simulator: User's Guide*. Organizational Learning Center, Massachusetts Institute of Technology. Cambridge, MA. April, 1993.

Oliva, R. 1993b. *Service Quality-Service Capacity Interactions: Framework for a Dynamic Theory*. Systems Dynamics Group, Massachusetts Institute of Technology. Cambridge, MA. November, 1993. D-4371-2.

Oliva, R. 1996. *A Dynamic Theory of Service Delivery: Implications for Service Quality*. PhD Thesis, Sloan School of Management, Massachusetts Institute of Technology.

Oreskes, N., K. Shrader-Frechette and K. Belitz. 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, **263**, 641-646.

Powell, M.J.D. 1969. A method for non-linear constraints in minimization problems. In R. Fletcher (Ed.), *Optimization*. (pp. 283-293). New York: Academic Press.

Senge, P.M. 1977. Statistical estimation of feedback models. *Simulation*, **28** (June), 177-184.

Senge, P.M. 1990. Catalyzing Systems Thinking within Organizations. In F. Masaryk (Ed.), *Advances in Organizational Development*. (pp. 197-246). Norwood, NJ: Ablex.

Senge, P.M. and R. Oliva. 1993. Developing a Theory of Service Quality/Service Capacity Interaction. In E. Zepeda and J.A.D. Machuca (Ed.), *1993 International SD Conference*, (pp. 476-485). Cancún, México.

Senge, P.M. and J.D. Sterman. 1992. Systems Thinking and Organizational Learning: Acting Locally and Thinking Globally in the Organization of the Future. *European Journal of Operational Research*, **59** (1), 137-150.

Sterman, J.D. 1984. Appropriate Summary Statistics for Evaluating the Historical Fir of System Dynamics Models. *Dynamica*, **10** (Winter), 51-66.

Ventana Systems. 1995. *Vensim 1.62 Reference Manual*. Belmont, MA: Ventana Systems, Inc.

Zeithaml, V.A., A. Parasuraman and L.L. Berry. 1990. *Delivering Quality Service: Balancing Customer Perceptions and Expectations*. New York: The Free Press.