

Interviewing as a Strategy for the Assessment of System Dynamics Models

Luis Felipe Luna-Reyes,¹ Vedat G. Diker,² and Deborah Lines Andersen³

¹ *Universidad de las Américas, Puebla, School of Business, NE221J, Santa Catarina Mártir, Cholula, Puebla, México 72820, Phone: +52 (222) 229-2000 ext. 4536, Fax: +52 (222) 229-2726 email: luisf.luna@udlap.mx*

² *University of Maryland, College of Information Studies, Room 4105 Hornbake, College Park MD 20742, Phone/VM: +1 (301) 405-9814, Fax: +1 (301) 314-9145, email: vdiker@umd.edu*

³ *University at Albany, SUNY, School of Information Science and Policy, Draper 140C, 135 Western Avenue, Albany, NY 12222 Phone: +1 (518) 442-5122, Fax: +1 (518) 442-5367 email: dla@albany.edu*

Abstract

System dynamics requires the intense use of qualitative data and human judgment in all stages of model development. Most approaches to the formal inclusion of qualitative data have been developed with the purposes of knowledge elicitation during the conceptualization or formulation stages of model development. Although the importance of using expert judgment to assess the validity of system dynamics models is well recognized, the development of approaches to use this kind of judgment is not well developed. In recent years, efforts to develop tools to assess the validity of system dynamics models by interviewing experts have been explored in some doctoral work. This paper reviews the basic concepts of model validation, and explores the use of interviews as a research and knowledge-acquisition technique. Finally, it documents and compares four applications of interviewing as a tool to assess system dynamics models, ending with recommendations for both the practitioner and researcher.

Introduction

Qualitative data is recognized as the main source of information to develop system dynamics models (Forrester, 1992). Based on the fact that qualitative data and judgments are actually much more used by managers in the development of strategy and decision-making processes, Wolstenholme (1999) calls for the development of methods and skills to engage qualitative thinkers in the whole process of model development. From the methodological point of view, recent efforts also call for the development of protocols to

promote the use of qualitative data gathering and analysis techniques during the modeling process (Luna-Reyes and Andersen, 2003).

Although protocols and approaches to using qualitative data in model development have been created and tested, most of them focus on the conceptualization and formulation stages of model development (Vennix *et al.*, 1992; Ford and Sterman, 1997; Lee *et al.*, 1998). Although the importance of the use of qualitative data and human judgment during the validation of system dynamics models is also recognized in the literature (Forrester and Senge, 1980), there are fewer examples of the use of these data in the later stages of the modeling process.

Interviewing is one of the most widely used methods of gathering qualitative data in social research (Fontana and Frey, 2003), and some recent doctoral work has been experimenting with the use of interviews in the validation stage of model development (Black, 2002; Rich, 2002; Diker, 2003; Luna-Reyes, 2004). This paper documents and compares these efforts.

After this brief introduction, the paper is organized in five interrelated sections. The first reviews the concept of model validation as understood by the system dynamics community. The second section reviews the principles to develop interviews as social research tools. The following section makes a brief description of four efforts to use interviews in the validation of system dynamics models. The fourth section contains a comparison of these four approaches. The paper ends with a reflection on the four experiences, and suggestions for further research in the area.

Considerations on Model Validation

Validation is an important issue in the development of system dynamics models as it is in any other kind of model (Barlas and Carpenter, 1990; Barlas, 1996). The problem is complex given that validation of a system dynamics model is not the last step in the modeling process, but it is intertwined along the whole process (Richardson and Pugh, 1981). From the system dynamics point of view, validate a model is impossible, given that “all models are wrong” (Sterman, 2000). Rather, it is common to describe the

process as “building confidence” in the model relative to some specific purpose (Richardson and Pugh, 1981; Sterman, 2000). As posed by Forrester and Senge (1980), the validation of a model has as its main purpose to attain “transferred confidence in a model’s soundness and usefulness as a policy tool” (211). Tying the validation process to the main purpose for which the model was created promotes a validation method that uses a series of semi-formal processes involving a “social conversation, rather than objective confrontation [with reality]” (Barlas and Carpenter, 1990:163).

This view on model validation has created several debates about the formality of system dynamics as a scientific method of inquiry.¹ Barlas and Carpenter (1990) discussed this controversy in terms of the philosophy of science, concluding that “the views of system dynamicists on validation parallel the relativist philosophy of science” (162). That is to say, from the traditional logical empiricist point of view, the system dynamics method does not fulfill the criteria of “good science,” but it adheres to the practices of theory confirmation followed by the contemporary relativist point of view.

Consistent with the relativist philosophy of science, the validation process in system dynamics considers the use of many different tests to promote the conversation about the adequacy and confidence of the model in terms of its structure and behavior (Forrester and Senge, 1980; Richardson and Pugh, 1981; Sterman, 2000). Although many of these tests point to the need for collecting and analyzing qualitative expert judgment or published literature to assess the adequacy of model structure and behavior, most of the literature focuses on quantitative methods to assess model validity (Sterman, 1984; Barlas, 1989; Barlas, 1990; Balci, 1994; Kleijnen, 1995; Oliva, 2003; Oliva, 2004). From the academic point of view, this lack of balance could be explained because the logical empiricist paradigm is still dominant in many fields that use system dynamics as a research method. From the consultant point of view, client expectations constitute an important source of pressure to pay special attention to formal, mathematical tools to assess model validity (Homer, 1997; Coyle and Exelby, 2000).

In this way, this paper contributes to the literature on model validation by providing tools and methods for using interviews to collect and analyze qualitative expert judgment in

the assessment of system dynamics models. The interview seems an appropriate tool to promote social conversation about the adequacy of a model for a given purpose. Our goal is not to suggest that the use of a qualitative approach is superior to the use of quantitative ones, but to offer a complementary tool. The use of interviews becomes more important in system dynamics projects where modeler and clients do not have data series, and the modeling effort relies on the use of qualitative patterns of behavior. This is not uncommon in system dynamics projects as it is pointed out by Richardson and Pugh,

one [the modeler] is often faced with a dynamic problem in which a key variable is not traditionally quantified or tabulated. It is even more likely, however, that the modeler or the client knows the dynamic behavior of interest without referring to data (1981:19).

In the following section, we describe the different types and methods to conduct interviews, followed by some practical guidelines and current practices to analyze interview data.

Interviews as a Social Research Method

Interviewing is a time-honored method of collecting data for research. Historians use this technique to create oral histories that will be published, in their entirety, for subsequent analysis and enjoyment (Bryman, 2004:541). Social science researchers use interviews to gather data that they will subsequently analyze and use to generate, confirm or disconfirm research hypotheses.

There is sometimes a fine line between survey research and interview research. Most of the time one thinks of survey research as being done at a distance, where the researcher and the respondent do not have face-to-face contact with each other. Respondents fill out surveys and return them (mail, email, web-based, or a drop box) back to the researcher. Most often surveys contain specific questions that require short answers, a check in a box, or rating on a scale, with an occasional space to write several sentences reflecting on a particular topic. There are instances of telephone survey research in which the respondent and the researcher do have oral but not visual contact with each other and the

respondent answers questions verbally while the researcher checks off the answers on a paper survey or computer screen. Survey research often leads to highly quantifiable data.

Interview research is most often face-to-face, with extended opportunities for respondents to expand their answers. Whereas survey research usually involves a printed/written response, interview research normally involves an oral response that the researcher captures through notes, tape recording, or both. There is a lot of messy data involved in interview research. Respondents are free to elaborate upon their answers, telling stories to illustrate their points. Researchers can, in turn, ask for elaborations or clarifications on what has been said. These data might explain understandings, processes, understandings, feedback mechanisms or causal relationship. Although the data can be quantified, it takes some work on the part of the researcher to actually do this. Mode (oral) and richness of response (very detailed) are two important characteristics of the interviewing method.

Interviewing Methods

After developing a statement of the research problem, one of the next most critical methodological issues is deciding upon an appropriate population of respondents who have information that the researcher can uncover. Next, the researcher must decide what method of data collection will be best. When quantities of easily quantifiable data are needed, survey aimed at the entire population, or a sample of that population, will yield many data points from a wide range of respondents. When, however, the researcher needs rich stories from respondents, then interviews are often the best research methodology. As stated above, interviews lend themselves to description of processes, understandings, causality, and feedback. They are labor- and time-intensive for the researcher but can yield subtleties of understanding impossible with paper or web-based surveys.

Types of Interviews

Interviews can take many forms. Structured interviews in many ways look like surveys. The researcher comes prepared with a very formal list of questions. The interview script might include boxes where the interviewer can check responses. Because this is an interview, the respondent has the ability to elaborate upon answers or to question the

researcher before indicating a response. The questions asked of each respondent in the sample will be much the same, following the structured interview script (Babbie, 1992:269-275).

Unstructured interviews, sometimes referred to as “intensive interviewing,” or “journalistic interviewing” (Spradley, 1979:58-68; O’Sullivan and Rassel, 1989:190) leave the questions and direction of the interview to the researcher and the respondent. One might expect that each respondent would have a very different interview, based upon his or her experience and the direction that the interview takes.

Semi-structured interviews fall somewhere between the two extremes above. The researcher might start with a series of structured questions and then move to a more journalistic approach, letting the interview proceed as a conversation.

Within this range of structure through unstructured interview forms there are other ways that interviews might vary. Sometimes it is more appropriate to have a group interview in which individuals build upon each other’s responses and develop rich data based upon group feedback. These group interviews (e.g., group model-building sessions in system dynamics) are often referred to as “focus groups.” Their main drawback is that one or two individuals in the group can anchor the thoughts of the rest. Individual interviews avoid this anchoring. There are strengths and limitations to both approaches.

Another variable in interviewing is the choice of technology for delivery. One might conduct an “interview” through the mail or preferably through email in which the researcher and respondent would carry on a “conversation” either in real time (e.g., instant messenger mode) or off-line with the two responding over several hours, days, or weeks. This method eliminates face-to-face dialog and reading of body language that can either be strengths or limitations of the interview process. Using email also means that the entire “conversation” is recorded in digital format, making transcription of tapes unnecessary, unlike telephone interviewing in which the researcher must take notes and/or record the conversation and have it transcribed. Researchers should note the problems with asynchronous interviewing in that both interviewer and respondent can lose the thread of the “conversation” or become weary of the extended dialog.

Video conferencing, recording both the video and audio aspects of the interview, is yet another method of capturing the data from interviews when it is not practical for the researcher and/or respondent to travel for a face-to-face meeting. Especially when there is video at both ends of the interview, this is a strong substitute for face-to-face communication when body language and nuance indicated by it are important to the research effort (Powell and Connaway, 2004:149, 155-157).

Practical Guidelines

There are a variety of practical matters that make interviewing more or less difficult. The following is a list of issues to keep in mind before starting the interview process.

- There are advantages to interviewing a respondent on his or her home turf. Respondents are normally more relaxed and can refer to materials they have at hand. This also means that the researcher, rather than the respondent, travels to the site—time-expensive for the researcher but also advantageous if one wants the respondent to be fresh and at ease.
- Interviewing is hard work. It is best to allow time beforehand to review questions and find the interview site, and time afterwards to take notes and think about the interview. One might expect to spend one outside hour for every one hour of interview (and an interview of more than one hour is probably too long). Two or three interviews a day are probably the most that a single interviewer can manage without the interviews blurring into each other.
- The researcher needs a data collection instrument, commonly referred to as a “script.” The script can have very detailed questions—looking much like a survey, or a series of very open-ended guidelines. There should be one script for each respondent so that the researcher can also fill in name, date, time, place, notes, and the number of the tape that goes along with the interview. One can also fill in the transcript number once the data have been transcribed from tape to text. Once the researcher has conducted many interviews these details will be critical. Sometimes respondents will want to know the nature of the interview beforehand.

The script easily provides this information and can be sent in advance of the session.

- The script should be pilot tested before actual administration to the respondent pool. Pilot testing will clear up any inconsistencies, badly worded questions, or omissions of critical questions. Order of discussion is very important in an interview. Working through an actual interview allows the researcher to test the flow of conversation and information.
- There are important ethical concerns involved in any project with human subjects. In general, subjects need to know the purpose of the research, why they were selected, the procedure that the researcher will follow, how long the interview will take, how the data will be used and disposed of, who will see the data, and the costs and benefits to themselves. Additionally, the researcher must obtain consent from the respondent for the interview and, when appropriate, for recording. Finally, human subjects review requirements in the United States require that participation is voluntary—a participant may choose to stop at any point during an interview (Office of Human Research Protections).
- With small studies it is sometimes enough just to take notes and get data for analysis. Usually, one must go one step farther and transcribe the audiotapes in order to properly analyze them. It takes approximately 30 double-spaced pages for every hour of interview taped. There are individuals who make a living doing transcription. It is usually less expensive to pay them than to try and use transcription equipment oneself. It is incumbent upon the researcher to check the reliability of the transcription, listening to the tapes with the manuscript in hand. The transcriber will be less familiar with the subject matter than the researcher and might miss words or phrases that are specific to the issues addressed in the interview.

Analyzing Interview Data

Depending upon the original research problem, there are a variety of methods the researcher might use in coding data. As mentioned above, sometimes it is just enough to listen to the tapes again, noting a particular process or explanation that is central to understanding of the problem. Usually, however, the researcher must use more formal coding methods in order to make sense of the data contained in interview transcripts (Bryman, 2004:146-150).

In formal content analysis the research develops a set of codes—words or phrases that he wants to find throughout all the interview transcripts. Although original content analysis was done by hand, with the researcher reading the manuscripts and marking target expressions, it is now common for researchers to load a word-processed file and analyze it with software designed for content analysis. These software packages will pull out specified instances of words, phrases or sentences and can be set to display whole sentences or paragraphs that contain the elements desired. (NUD*IST and ATLAS.ti are two of the most widely used software tools.)² Not only is this method much faster than hand coding, but it eliminates the problems of inherent in human error, or inter-coder reliability when more than one person is doing the coding.

Four Experiences on the Use of Interviews to Assess System Dynamics Models

After describing the main methods and techniques for conducting interviews, this section of the paper is oriented to present four illustrations about the use of interviews in the validation of system dynamics models. Examples are extracted from recently finished doctoral dissertations. Each example includes a description of both the gathering and analysis techniques used. Although all the projects presented in this section used a variety of techniques to assess the validity of the model during the whole process, the description of this section focuses only in the interview component of the strategy.

Exploring the Dynamics of Collaboration across Interdepartmental Boundaries

Gathering Data

Investigating interdepartmental interactions and collaboration during project development, Black (2002) developed a model-based theory of collaboration relying in data gathered by intensive field research. The site was a medium-size company that manufactures motor vehicles. The guiding question of her research project was

Given that expertise is distributed throughout an organization, how do the nature and timing of interdepartmental activities [between design engineers and manufacturing and assembling personnel] bring that expertise to bear effectively on product development efforts? (107).

Black's data gathering approach for model development consisted of three different stages. The first two stages focused in the general process of product development inside the company, and the third stage was the follow up of one particular project called "Hook". The model developed by her was based upon this project.

To assess model validity, Black interviewed participants in the Hook project from both the design, and manufacturing and assembling departments. She used unstructured interviews based on causal diagrams showing main pieces of model structure, together with a range of simulations. She showed each picture to the interviewee asking for the face validity of both the causal relationships and the patterns of behavior produced by the model.

As an additional activity, she reviewed documentation and interviewed managers of other 5 projects for comparison purposes. Although she does not include these interviews as part of the validation strategy, we consider that looking for additional cases that could help to look for commonalities or to identify alternative theories to explain problems in product development constitutes an important activity to increase confidence in the adequacy or to validate the model.

Analyzing Data

Unfortunately, Black does not describe the way in which she analyzed the data gathered through her confirmatory interviews used to validate the model. By the review of her dissertation, it looks that she engaged in the process of social conversation to analyze the face validity of causal relationships and behavior, and no major questions were raised during those conversations.

Exploring the Sustainability of Knowledge Management Projects

Gathering Data

Rich (2002) used a system dynamics approach to explore how knowledge management programs in a company affect the overall company environment; and how those programs are in turn affected by the overall environment. He developed a system dynamics model of knowledge management activities within a firm. The model was based on field studies carried out at two consulting firms which had implemented knowledge management programs.

Rich tested his model by interviewing a number of domain experts from one of the companies who participated in the model development phase. Although there were a total of 11 subjects who participated in the model development phase, only four were available for the second phase of the interviews, where the actual testing of the model took place. In that sense, sampling was purposive and limited to the entirety of a small population. Rich implies that group interviews would be more desirable; however, he had to do individual interview, since the logistical barriers did not allow for a gathering of the subjects. The interviews were carried out over the telephone. The subjects received an introductory “booklet” prior to the interviews, which consisted of two portions: one for pre-interview orientation, and one to be used during the interview. The pre-interview portion included a simplified version of the model with six internal variables, four external inputs and two external outputs. This simplified version of the model involved three feedback loops, and was presented as a single causal loop diagram. The pre-interview portion of the booklet also included explanations about the assumption of the

model, the performance indicators, and five policy scenarios the outcomes of which were used as the basic context for the interviews.

During the interviews, the subjects speculated about the potential outcomes of the five policy scenarios and commented on the plausibility of simulated model behavior under each policy scenario. The subjects rated the behaviors of five performance indicators as “Plausible,” “Uncertain,” or “Not Plausible.” By doing this, Rich tried to surface the relevant mental models of the subjects, and have them evaluate the simulation model based on them.

The interviews were recorded on cassette tapes and later transcribed. Two coders coded the subjects’ evaluations of the indicators under each scenario. Wherever there was no explicit evaluation, the subject’s intent was coded.

Analyzing Data

While the small sample size did not allow for any kind of statistical or other quantitative analysis, it enabled Rich to portray a substantial portion of the interview data directly. Rich tabulated the expert evaluations of the behaviors of the performance indicators based on the coded data. He also summarized the detailed expert comments wherever important insight or intricate differences in expert evaluations had emerged. He also interpreted the interview data in order to devise “indicated changes” to the model. The changes indicated were discussed in three headings: changes in model behavior, changes in model structure, and additional performance indicators. Specifically, there were four suggested changes related to model behavior, three changes to model structure, and two potential new performance indicators.

Exploring the Dynamics of Online Communities

Gathering Data

Diker’s dissertation (2003) explored the dynamics of growth in online communities that focus on developing and disseminating information artifacts such as open source software and instructional materials. Diker developed a system dynamics model of growth in a

hypothetical open source software development community, which he argued is an example of what he called “open online collaboration communities (OOCC),” or “open online content development communities (OOCDC).” He hypothesized that both open source software development communities and instructional materials development communities are OOCCs, and consequently they share common dynamics that govern their growth and decline. Based on this hypothesis, he tested the structure and the behavior of his model against the experiences and mental models of the leading members of an instructional materials development community. The community in question was a group of educators and researchers who work on applying systems thinking and system dynamics concepts to K-12 education. Diker referred to that group as “System Dynamics K-12 Community.” He used a purposive, snowball sample, and interviewed a total of 10 leading members of the system dynamics K-12 community. He used semi-structured, individual interviews. The interviews were carried out over the telephone, with the exception of one, which was carried out face-to-face upon the subject’s request. The interviews were recorded on cassette tapes. Prior to the interviews, Diker mailed the interviewees a packet containing a consent form, simplified feedback diagrams of his model, and empty charts to be used for plotting behavior-over-time graphs. The ways in which the diagrams and the charts were used during the interviews is explained below. The interviewees mailed back the consent forms, and plotted charts to Diker in postage-paid, self-addressed envelopes. The interviewees were told not to look at those materials prior to the interview with the exception of the consent form.

Diker tried to capture the interviewees’ observations and mental models about their community as they are pertinent to the structure and the behavior of his model. He started the interviews with questions that were aimed at having the interviewees think about the dynamics in their community. He then asked the interviewees’ to plot behavior-over-time graphs for some key variables. He asked the interviewees to choose the key variables to plot. He also prompted them to plot five key variables he defined, if one or more of those were not included in the interviewees’ choice of key variables. The interviewees plotted two types of behavior-over-time graphs; one based on their observations so far, and one based on their projections for the future. The second type involved several “futures” or “scenarios” in many cases. While a few of the interviewees completed drawing the

behavior-over-time graphs during the interviews, most of them asked for extra time to work on them after the interview.

After the portion about the behavior-over time graphs, Diker introduced a loop-by-loop description of a simplified version of his model to the interviewees. He used a series of seven sketches to unfold the simplified model one loop at a time. A total of five major loops were introduced. The final sketch involved 13 variables. A narrative was read to the interviewees as they observed the sketches. The interviewees were asked to stop the narrative whenever they observed an argument about a variable or a causal link that they think does not exist in their community. At that point the interviewee could comment on why they think the variable or the link was conceptualized wrong, and what they think would be a better conceptualization.

After all the loops were introduced and discussed, Diker explained to the interviewees how the given model defines the main policy problem within an OOCC. He then asked the interviewees to comment whether they observe a similar problem prevailing in their community. Diker also asked the interviewees about the existence or applicability of four policy options, which were conceptualized to remedy the main policy problem, within the context of their community. The policy options were introduced as unfolding loop diagrams, based on the simplified model initially introduced to the interviewees. The interviewees were finally asked which of the policy options or a combination there of, would work best in their community. At that stage, the interviewees could comment on the positive and negative consequences of each policy option.

Analyzing Data

Diker carried out the analysis of the interview data by reviewing the sound recordings of the interviews. Given the small sample size and the richness of the data, he concluded that summarizing each interviewee's comments on each loop and policy option would be a feasible way to analyze and report the interview data. He built a two-column table for each loop and policy option, where the first column denoted the interviewees and the second column contained the comments by the interviewees on the given loop or policy option. He also wrote a summary analysis of comments for each loop and policy option,

where he grouped similar comments and discussed what they imply about the model structure and the potential success of the policy options. Concluding, he argued that the overall analysis of comments implied limited changes for the system dynamics model, and provided insights about the applicability of the policy options within the context of the system dynamics K-12 community.

Exploring Collaboration in Interorganizational Information-Technology (IT) Projects in Government

Gathering Data

Interested in the collaboration and trust development processes in the development of interorganizational IT projects, Luna-Reyes (2004) conducted a model-based case study analyzing two IT projects in New York State.³ The first project consisted on the development of a prototype to better manage homeless services through the State known as the Homeless Information Management System (HIMS). The second project consisted in a contact repository and document management system developed at the New York State Office of the State Comptroller as a tool to manage services to State local governments, the Multi-purpose Access for Customer Relations and Operational Support (MACROS). The approach consisted in developing a simulation model based on the HIMS case that was assessed comparing its main assumptions with the experience of a group of people from the MACROS project.

The validation strategy used by Luna-Reyes consisted in two main components. The first component of the strategy consisted in three interviews conducted with four participants in the HIMS project (all of them were interviewed during the conceptualization and formulation of the model, two of them were interviewed in group) to assess the structural and behavioral correspondence of the model with their experiences. The interview was facilitated by a high-level picture of the theory, and 23 behaviors over time of key accumulations in it. The description of the high-level picture of the model included the main logic and basic assumptions included in it, and the descriptions of each of the behaviors included the main structural components associated with each of them in form of a story. In many cases, interviewees asked for additional information about model

structure, which was provided by the use of simplified graphs of structures in the model closely related to each behavior. The interviewees had the opportunity to clarify the story, and to make comments about the correspondence of each of the behaviors to his/her experience in the project. Considering that discussing 23 behaviors during a one-hour interview was an aggressive objective, the behaviors were presented in different order to get at least two assessments for each of them. Fifty seven out of the 69 possible data points (3 interviews, 23 behaviors discussed in each), accomplishing the goal of having at least 2 assessments for each behavior. Each interview was recorded using both, notes and a tape recorder.

The second component of the assessment strategy consisted in the assessment of the applicability of some key model assumptions in a second collaboration experience. To accomplish this objective, Luna-Reyes interviewed six participants in the MACROS project. The interviewees collaborate in teams developing work related to the development and implementation of the MACROS system across the NYS Office of the State Comptroller. The interview consisted of the assessment of 16 statements containing some key model assumptions, organized around five small structural pictures. Each participant was asked to assess the statements on a scale from 1 to 5, where 1 corresponded to absolutely false and 5 corresponded to absolutely true. Respondents were also asked for examples or counterexamples for each statement according to their experience collaborating in the MACROS project. Again, each interview was documented with notes and a tape recorder.

Although many structural components were introduced in the interviews HIMS participants, the interviews were oriented to discuss the behavior of the model. On the other hand, the interviews with the MACROS participants had a focus on structural components and assumptions of the model, to assess the transferability of the structural assumptions to other collaborative experiences.

Analyzing Data

Interview data analysis can also be differentiated as following two different strategies, one for each kind of assessment interview. In both cases, each tape was reviewed several times to complete the notes taken during the interview.

For the interviews with HIMS participants, each of the answers was coded to reflect the correspondence of model behavior to interviewee experience. The coding scheme also included the fact that the mismatches between model output and behavior corresponded to problems with values of parameters or to structural problems in the theory. In this way, five codes were used to classify each response (see Table 1). Additionally, all interviewees' comments were summarized, documenting parameter changes that emerged from each comment. Coded answers were used to discuss model adequacy.

Table 1. Coding scheme for the interviews with HIMS participants. Reproduced from Luna-Reyes (2004:291)

No.	Code	Description
1	+Nothing	The qualitative pattern of behavior, and the timing of the behavior (periods of growing, for example) correspond to the perception of the HIMS case. No additional comment.
2	+Story	The qualitative pattern of behavior, and the timing of the behavior (periods of growing, for example) correspond to the perception of the HIMS case. Some clarifications to the story were made.
3	-Fixed	The pattern corresponded, but the intensity, initial or final values did not correspond to the experience in the HIMS project. The difference between the model behavior and the perception of the interviewee was corrected by making parameter changes.
4	-Not Fixed	The pattern did not correspond, intensity, initial or final values did not correspond to the experience in the HIMS project. The difference requires structural changes in the model
5	Missing	No response for the behavior.

The second set of interviews was much more structured, and as a result much simpler to analyze (every structural statement had an assessment in a 1 to 5 scale). Luna-Reyes obtained descriptive statistics for each statement, correcting for possible response bias. As a result, he classified all statements in terms of the level of their correspondence with interviewees' experience at the MACROS project. Finally, he discussed the implications

of each mismatch in terms of the examples and counterexamples used by respondents, which were transcribed from the interviews' tapes.

Comparison and Assessment of the four approaches

A common characteristic of the four model-based research projects presented in the previous section is that the four of them made intensive use of qualitative data during the whole modeling process. Moreover, the four projects include the use of interviews during the validation stage of the model in an explicit and formal way. However, the approaches are different in several dimensions (see Table 2). The differences among the four projects constitute a good illustration of the possibilities and alternatives when using interviews.

Table 2 – Comparative view of the four approaches.

Characteristic	Black (2002)	Rich (2002)	Diker (2003)	Luna-Reyes (2004)
Who was interviewed?	Participants in the case modeled	Sample from a second case	Sample from a second case	From the case modeled, and from a second case
Technology of delivery	Face-to-face	Telephone (using a mailed booklet)	Telephone and one face-to-face (using a mailed booklet)	Face-to-face
Type of questions	Unstructured	Semi-structured	Semi-structured, starting with open questions before presenting model	Unstructured, and structured
How behavior was presented	Individual graphs over time	Scenarios with a selection of key indicators	Behavior embedded in stories. Captures new behaviors over time	Individual graphs over time
How structure was presented	Diagrams	Stories describing feedback	Diagrams supported by feedback stories	Diagrams and written statements
Recording technique	Notes and tape	Notes and tape	Notes and tape	Notes and tape
Data processing	None	Transcribing tapes	Notes and reviewing tapes	Notes and reviewing tapes
Data analysis	Looking for face validity	Coding (2 coders) ⁴	Summarizing and loose coding	Coding/ Descriptive statistics

Another important commonality of each of these four examples is that the main purpose of the modeling was oriented to increase our general understanding of a dynamic phenomenon, and not only to offer insight about a particular problematic situation. In this way, all of them offer to the reader some reflections about the validity of the models as general theories, making comments about limitations in data gathering and analysis. For example, Black (2002) discuss as a limitation of the approach the feasible existence of a retrospective bias, given that it is hard for respondents to interviews to remember pieces of data relevant for her theory such as “what was *unknown* at a particular time”, creating a tendency to “forget the assumptions that guided particular decisions” (122). Extensive thinking and reflection about the HIMS project among participants increased the reliability of Luna-Reyes’ (2004) theory (i.e. there is high consensus about the main components of the story). However, the same fact limits the validity of the theory because it also limits the access to alternative stories and causal explanations of the project success.

Although all of them also discuss the activities or “safeguards” used to deal with these limitations, there is still much to learn about it.

Concluding Remarks

The purpose of this paper was to contribute to the literature on system dynamics model validation by providing concrete qualitative tools to be used during this stage of the modeling process. We pointed out that interviews constitute a good fit to promote the “social conversation” proper of system dynamics established validation practices. Our intention is not to suggest that these techniques should be used instead of other validation strategies, but in a complimentary way. However, given that the system dynamics method encourages modelers to use qualitative data in an intense manner, we think that the formal incorporation of such techniques is important to improve practice.

Besides providing with a comprehensive set of guidelines to conduct interviews, the four examples presented in the paper provide a variety of question formats and analysis techniques that could be used in the validation process. Reflections from the examples

suggest a risk of biases in the validation process that calls for the experimentation with different question formats in order to improve the effectiveness of the interview process.

Notes

¹ See Barlas and Carpenter (1990) and Barlas (1996) for citations in early debates about the validation procedures of system dynamics.

² Information about these software packages can be found at <http://www.qsr.com.au/> and <http://www.atlasti.de/>

³ Both cases used by Luna-Reyes (2004) were projects included in the “Knowledge Networking in the Public Sector” (KDI) research project developed at the Center for Technology in Government (CTG) at Albany, NY. The KDI project involved the longitudinal analysis of seven IT intensive innovations in the public sector. CTG’s work created a rich set of archival data for each case, mainly qualitative in nature.

⁴ To corroborate the reliability of coding process, Rich (2002) used a second coder, and calculating the inter-coder reliability coefficient developed by Cohen (1960)

References

- Babbie, E. (1992). The practice of social research. Belmont, CA, Wadsworth Publishing Company.
- Balci, O. (1994). "Validation, verification and testing techniques throughout the life cycle of a simulation study." Annals of Opns Research **53**: 121-173.
- Barlas, Y. (1989). "Multiple tests for validation of system dynamics type of simulation models." European J. of Operations Research **42**(1): 59-87.
- Barlas, Y. (1990). "An autocorrelation test for output validation." Simulation **55**(1): 7-16.
- Barlas, Y. (1996). "Formal aspects of model validity and validation in system dynamics." System Dynamics Review **12**(3): 183-210.
- Barlas, Y. and S. Carpenter (1990). "Philosophical roots of model validation: Two paradigms." System Dynamics Review **6**(2): 148-166.

- Black, L. J. (2002). Collaborating across boundaries: Theoretical, empirical, and simulated explorations. Sloan School of Management. Cambridge, MA, MIT.
- Bryman, A. (2004). Social research methods. Oxford, Oxford University Press.
- Cohen, J. (1960). "A coefficient of agreement for nominal scales." Educational and Psychological Measurement **20**(1): 37-46.
- Coyle, R. G. and D. Exelby (2000). "The validation of commercial system dynamics models." System Dynamics Review **16**(1): 27-41.
- Diker, V. G. (2003). Toward a dynamic feedback theory of open online collaboration communities. School of Information Science and Policy. Albany, NY, University at Albany, SUNY: 562.
- Fontana, A. and J. H. Frey (2003). The interview: From structured questions to negotiated text. Collecting and interpreting qualitative materials. N. K. Denzin and Y. S. Lincoln. Thousand Oaks, CA, Sage Publications: 61-106.
- Ford, D. N. and J. D. Sterman (1997). Expert knowledge elicitation to improve mental and formal models. 15th International System Dynamics Conference: "Systems Approach to Learning and Education into the 21st Century", Istanbul, Turkey, Bogazici University Printing Office.
- Forrester, J. W. (1992). "Policies, decisions, and information sources for modeling." European Journal of Operations Research **59**(1): 42-63.
- Forrester, J. W. and P. M. Senge (1980). Tests for building confidence in system dynamics models. System dynamics. A. A. Legasto, Jr. and e. al. New York, North-Holland. **14**: 209-228.
- Homer, J. B. (1997). "Structure, data, and compelling conclusions: Notes from the field." System Dynamics Review **13**(4): 293-309.
- Kleijnen, J. P. C. (1995). "Verification and validation of simulation models." European Journal of Operational Research **82**(1): 145-162.
- Lee, T. P., D. F. Andersen, G. P. Richardson, J. W. Rohrbaugh and A. Zagonel (1998). A judgement approach to estimating parameters in group model-building sessions: A case study of social welfare reform in dutchess county. 16th International Conference of the System Dynamics Society, Quebec '98, Quebec City, Canada, System Dynamics Society.
- Luna-Reyes, L. F. (2004). Collaboration, trust and knowledge sharing in information-technology-intensive projects in the public sector. School of Information Science and Policy. Albany, NY, University at Albany.

- Luna-Reyes, L. F. and D. L. Andersen (2003). "Collecting and analyzing qualitative data for system dynamics : Methods and models." System Dynamics Review **19**(4): 271-296.
- Office of Human Research Protections Us department of health and human services. <http://www.hhs.gov/ohrp/>. Last revised January 21, 2005
- Oliva, R. (2003). "Model calibration as a testing strategy for dynamic hypotheses." European Journal of Operational Research **151**(3): 525-568.
- Oliva, R. (2004). "Model structure analysis through graph theory : Partition heuristics and feedback structure decomposition." System Dynamics Review **20**(4): 313-336.
- O'Sullivan, E. and G. R. Rassel (1989). Research methods for public administration. New York, Longman.
- Powell, R. R. and L. S. Connaway (2004). Basic research methods for librarians. Westport, CT, Libraries Unlimited.
- Rich, E. (2002). Modeling the dynamics of organizational knowledge. School of Information Science and Policy. Albany, NY, University at Albany, SUNY: 351.
- Richardson, G. P. and A. L. Pugh, III (1981). Introduction to system dynamics modeling with dynamo. Cambridge MA, Productivity Press.
- Spradley, J. P. (1979). The ethnographic interview. Fort Worth, Harcourt Brace Jovanovich College Publishers.
- Sterman, J. D. (1984). "Appropriate summary statistics for evaluating the historical fit of system dynamics models." Dynamica **10**(2): 51-66.
- Sterman, J. D. (2000). Business dynamics : Systems thinking and modeling for a complex world. Boston, Irwin/McGraw-Hill.
- Vennix, J. A. M., D. F. Andersen, G. P. Richardson and J. W. Rohrbaugh (1992). "Model-building for group decision support and alternatives in knowledge elicitation." European Journal of Operations Research **59**(1): 28-41.
- Wolstenholme, E. F. (1999). "Qualitative vs. Quantitative modelling: The evolving balance." Journal of the Operational Research Society **50**(4): 422-428.