

SyM Bowl Update: Encouraging and Recognizing Proficiency in System Dynamics Modeling, Wayne Wakeland, Adjunct Professor of Systems Science, Portland State University, Portland, Oregon.

Abstract

The judging criteria and modeling process guidelines for the SyM Bowl modeling competition are briefly reviewed. Prize-winning student models from previous SyM Bowls are then discussed in order to show how the students and their teachers appear to have interpreted the criteria and process guidelines. The strengths and limitations of the selected student models, especially the basic model structure & logic, and the process they used to build the model, as evidenced in the reports they submitted, are then presented.

The paper then considers how the criteria and process guidelines may or may not have influenced the students, followed by a discussion on the strengths and weaknesses of the judging criteria and modeling process guidelines, including how they might be improved to assure that their influence on the student modeling process and model quality is as beneficial as possible.

Introduction

This paper discusses the SyM Bowl competition held annually in Portland, Oregon since 1996. It updates a previous paper by the author (Wakeland, 1998).

Since the earlier paper was published, two more SyM Bowl events have been held, and prior to each event the judging process was modified to some extent. The changes for 1999 were modest, with dimensional consistency being added to the list of criteria used to judge the model. For 2000, the changes were more substantial:

- The criteria were merged and condensed into 10 criteria, only 2 of which (vs. 4) are based on the poster sessions on the day of the event (see Appendix A)
- Judging of the student presentations was eliminated
- To a larger extent than in the past, the judging criteria emphasized results more than the process
- Rather than providing only overall scores for the model and for the paper, each judge provided scores for each of the individual criteria for each project (the avg. of these scores was provided to the student teams at the end of the event)
- To lessen to focus on competition, in addition to selecting 1st, 2nd, and 3rd place “winners,” all teams that achieved various “standards” would also be acknowledged. The concept of recognizing the achievement of standards was not fully implemented at the 2000 event due to miscommunication between the event organizers and the participating teachers (the notion of standards will be further refined and implemented at future events)
- The sample paper was revised somewhat to bring it into sync. with the updated criteria (Gallaher et al, 2000)

In future years, the process and criteria will continue to evolve in order to improve the quality of both the event itself and the modeling being done by the students. In a very real way, the SyM

Bowl participates in two key feedback loops, involving the organizers' efforts to improve SyM Bowl and the efforts by High School teachers to refine their approach to teaching System Dynamics (see Figure 1).

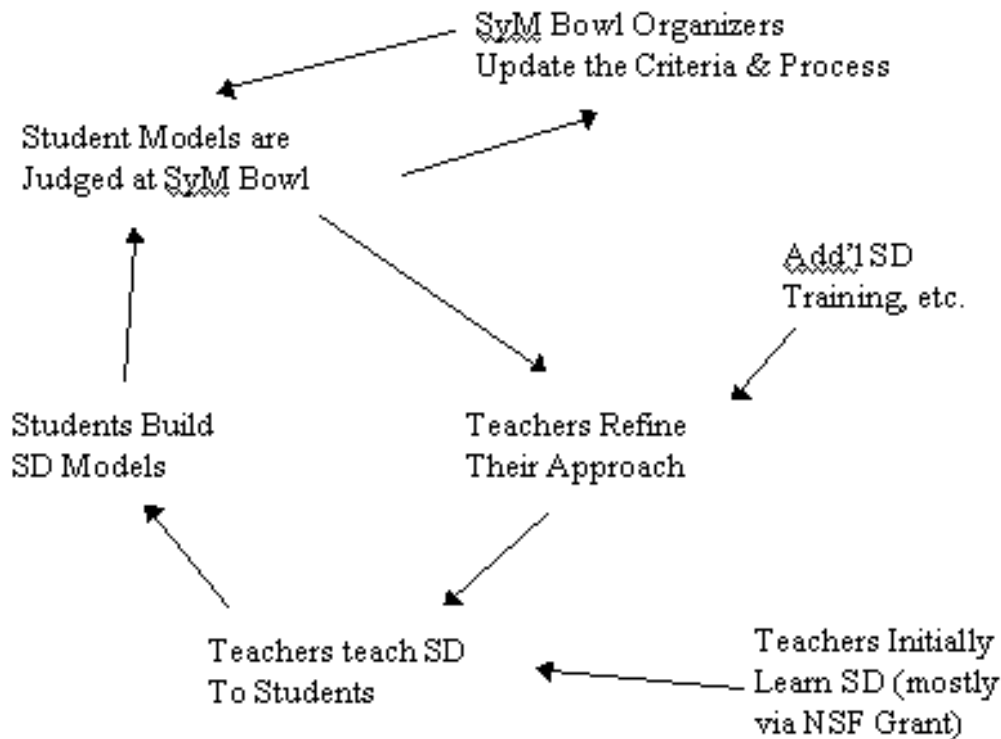


Figure 1: Possible SyM Bowl Feedback Loops

Exploratory Research on the Characteristics of Student System Dynamics Models

How are various characteristics of System Dynamics models built by High School students and submitted to SyM Bowl changing over time? Are they improving? If so, could this be due in part to the presence of these feedback loops? In order to begin to explore the answers to these and other questions, the author reviewed the top three to five projects from each SyM Bowl and made notes regarding the following aspects:

1. # of stocks
2. # of biflows
3. # of uniflows
4. # of non-constant converters
5. # of graphical functions
6. # of long FB loops (more than a self loop)
7. # of short FB loops (self loops consisting of a stock and one of its flows)
8. FB Analyzed?
9. Dimensional consistency?
10. Source of data for IV's & parms
11. Verification & validation
12. Reference Behavior Pattern?

- 13. Endogenous behavior?
- 14. Model Dynamics
- 15. Misc. Notes

The reason that the number of projects included herein varied from three to five for a given year is that the author looked for a “break point” that separated the top models from the rest of the submissions. The top three rated projects for each year were automatically included. One or two of the other finalists were included if their models were comparable to the top three in the opinion of the author (some projects were finalists because of the overall quality of their work, not just the quality of their model-- the focus of this paper).

For comparison purposes, 16 models built by graduate students were also reviewed. The resulting “dataset” is being published in a separate paper (Wakeland, 2000).

Using this dataset, the following measures were developed:

Table I: Measures for Model Comparison

MEASURE	DERIVATION
Model Size	Sum of Aspects 1- 4
FB Complexity	Aspect 6 plus Aspect 7 divided by 2
Model Development	Sum of Aspects 8 – 10 (each of which is “scored” from 0 to 2 or 0 to 3)
Model Testing	Sum of Aspects 11 & 12 (ditto)
Behavioral Complexity	Sum of Aspects 13 & 14 (ditto)

Admittedly, this scoring approach was highly subjective and was done by one person rather than by a panel of experts as would be preferable. The author felt that this approach was acceptable for an exploratory effort with an emphasis on comparison rather than objective measurement on an absolute scale.

The size measure was relatively straightforward. For FB complexity, it seemed reasonable to weight self-loops less than more complex loops, although, certainly, the weight chosen (.5) is arbitrary.

Regarding Model Development, which includes FB loop analysis, dimensional consistency and source of data for parameters and initial values, the intent is to indicate the degree to which the modeler has done two things: 1) *analyzed* their model and 2) supported their parameter values through research (either in the literature, via expert opinion, or by taking their own measurements).

Model testing combines the degree to which the model has been tested, and how completely and objectively the a priori reference behavior was established.

Behavioral complexity combines the degree to which the behavior is endogenously created, and how complex the behavior is...from simple linear to complex oscillations and complex overshoot/collapse/rebound.

The measures are plotted in Figures 2 through 6, with time as the independent variable to show how the scores have changed from year to year:

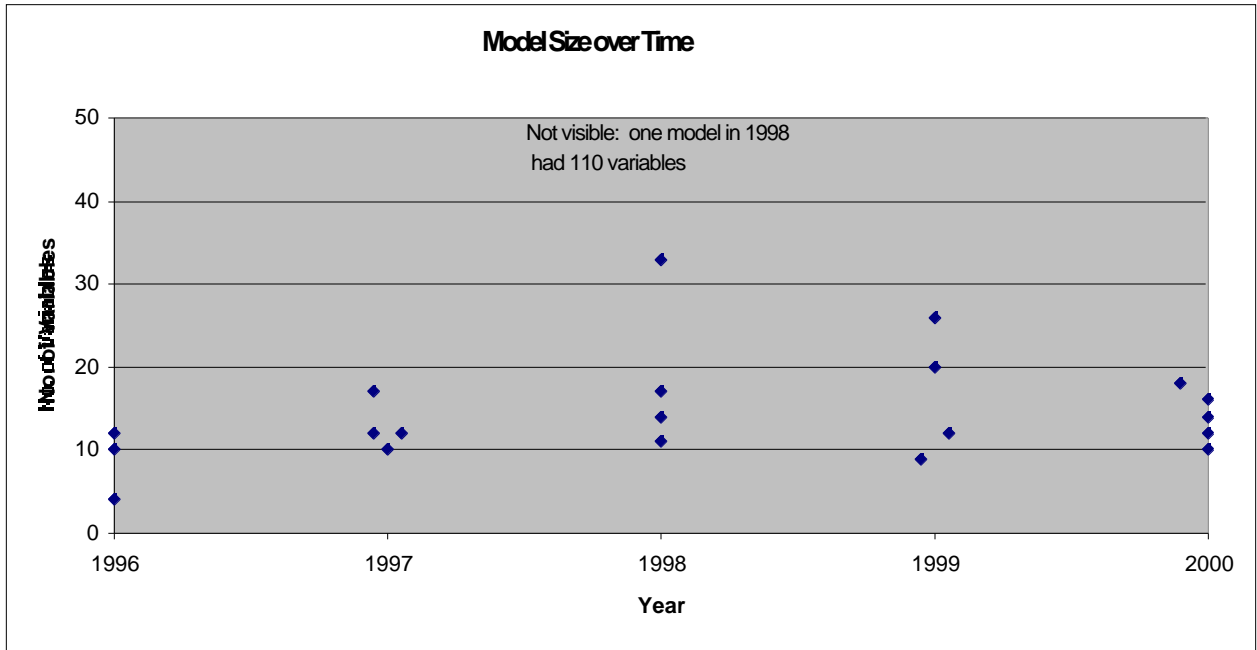


Figure 2: Model Size Measure

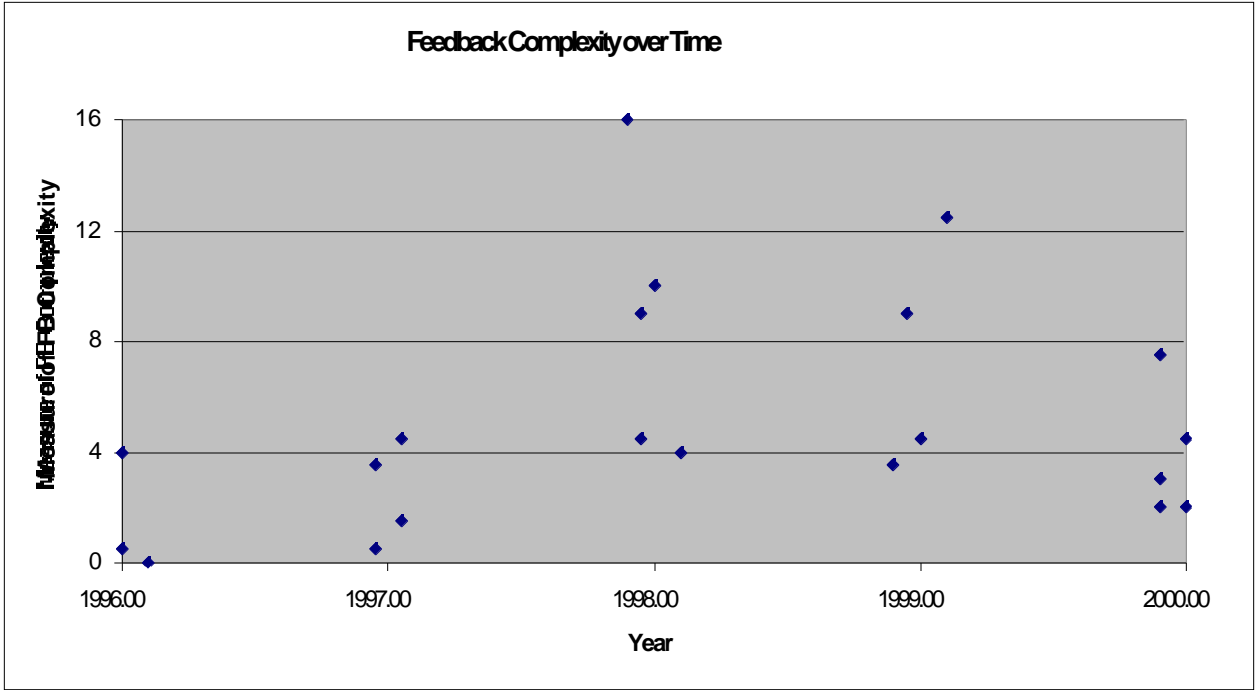


Figure 3: Model Complexity Measure

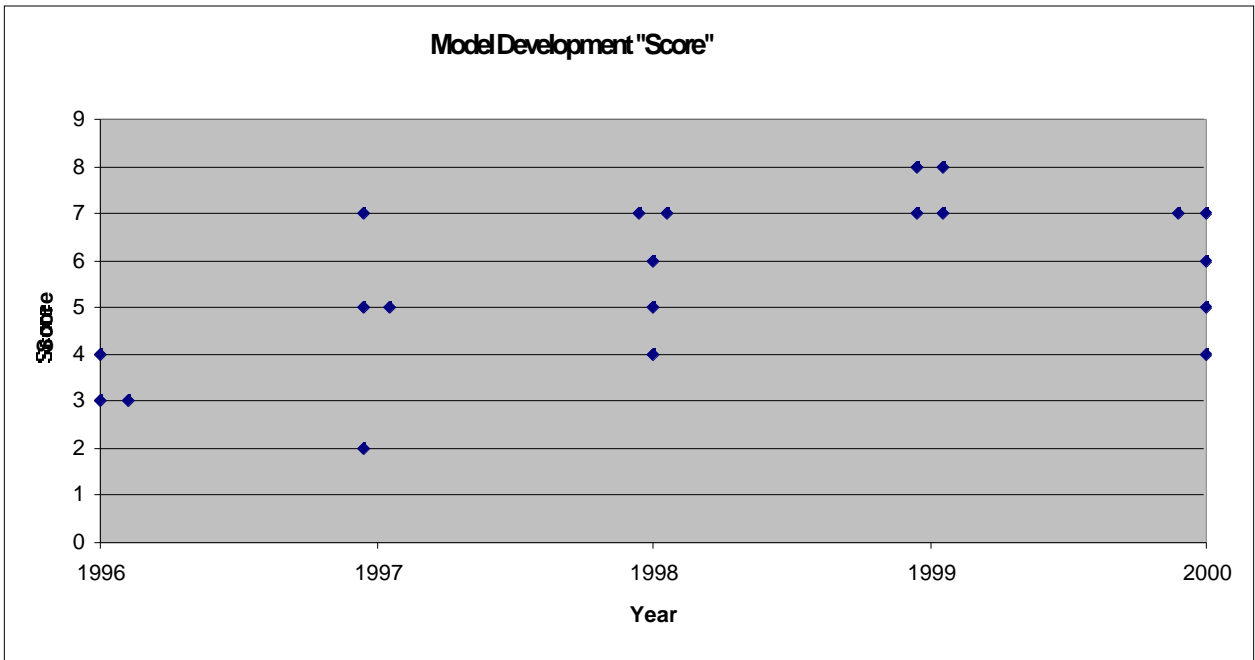


Figure 4: Model Development Measure

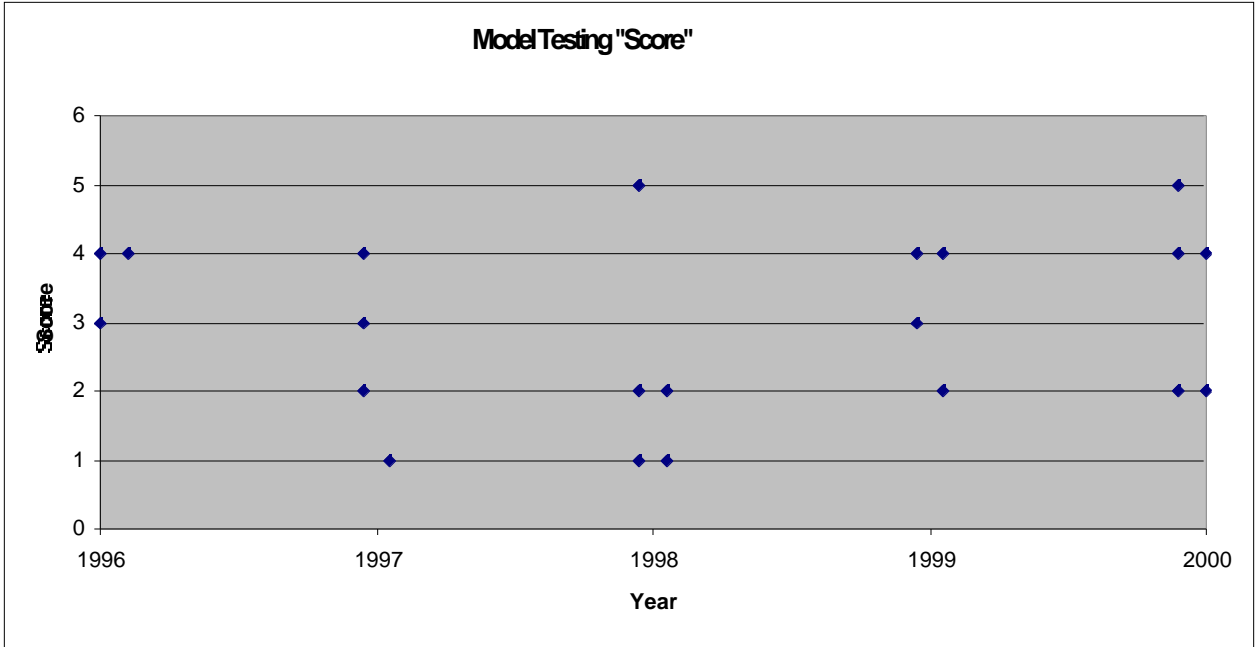


Figure 5: Model Testing Measure

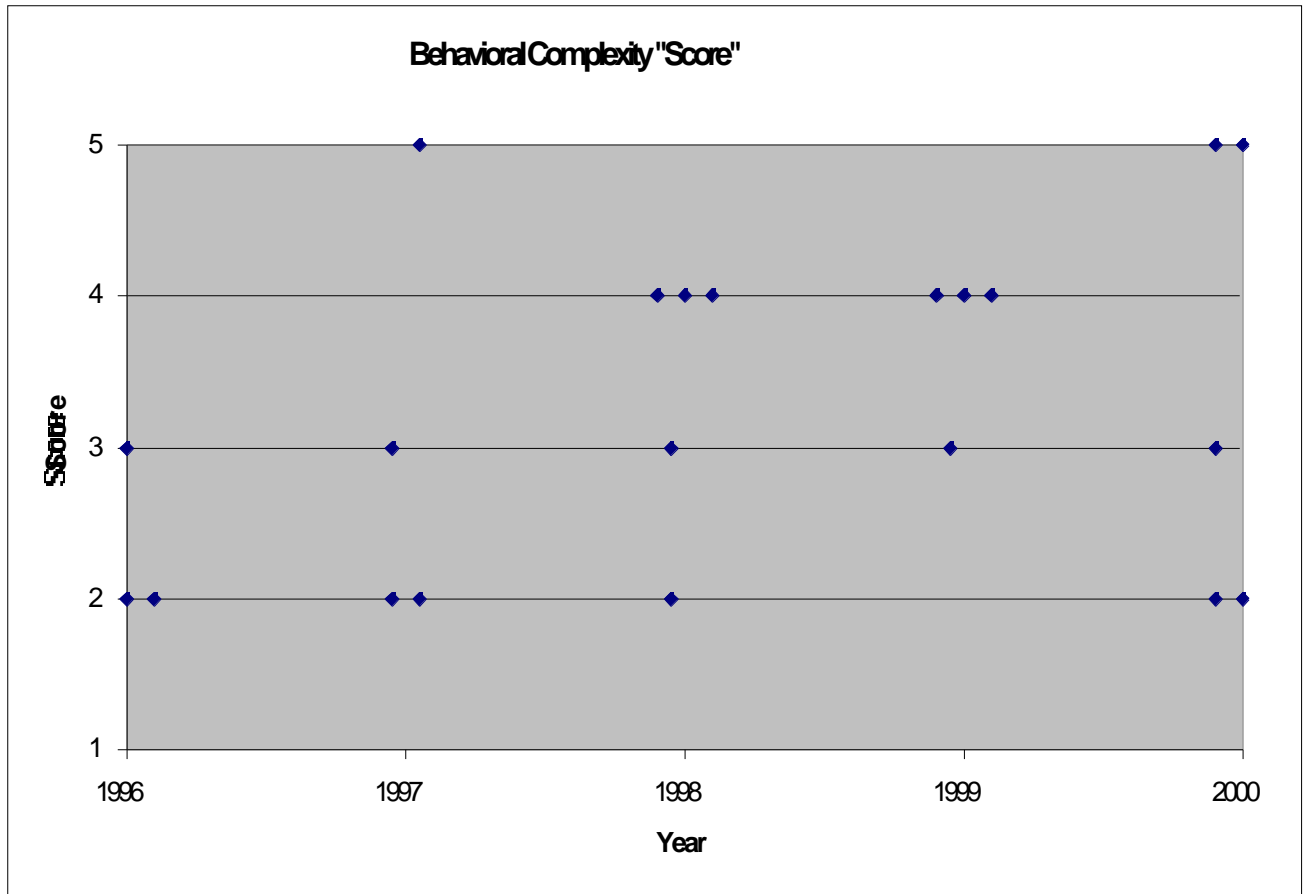


Figure 6: Behavioral Complexity Measure

Table II summarizes the mean values for each measure by year and overall. Also shown are the results of applying the same measures to a number of models built by graduate students.

Table II: Mean Values of Model Measures, over Time

Max. Score	MEASURE	1996	1997	1998	1999	2000	Overall Avg.	Grad. Students
	Mean Model Size	9	13	37	17	14	19	16
	Mean FB Complexity	1.5	2.5	8.7	7.5	3.8	5.1	6.3
8	Mean Model Development Score	3.3	4.8	5.8	7.5	5.8	5.6	3.7
6	Mean Model Testing Score	3.7	2.5	2.2	3.3	3.4	3.0	3.9
5	Model Behavioral Complexity Score	2.3	3	3.4	3.8	3.4	3.2	3.8
	Sample Size	3	4	5	4	5	21	16

What do these numbers imply, if anything? Many of the measures increase between 1996 and 1998 or 1999. Model Testing is the exception. Nearly every measure declined in 2000. Was

this change influenced by the change in judging criteria? This question is very difficult to answer. The SyM Bowl organizers want to believe that the revised criteria represent an improvement, but these results must, necessarily, raise at least some doubt regarding this belief.

On the other hand, the sample sizes are very small, and the scoring process was the work of one individual and highly subjective at that. It would clearly be inappropriate to infer anything of significance from this “data.” Still, the information is intriguing.

There is no reason to value larger, more complex models over smaller, less complex models. In fact, many would argue with sound reasons for just the opposite. In addition, complex behavior is not necessarily preferred to less complex behavior—except when it is indicated by the real world situation. This leaves only Model Development and Model Testing as indicative of actual “progress” over time. Development increased steadily from 1996 to 1999 and then returned to 1998 levels in 2000. Testing declined between 1996 and 1998 and then increased from 1998 to 2000. Since model size and FB complexity increase dramatically between 1996 and 1998, one could speculate that attention had shifted from thoroughly testing simple models to building more complex models with more complex behavior. If this is true, then this “trend” appears to have reversed in 1999, possibly allowing more time for model testing.

For curiosity’s sake, the final column of Table II shows the results of reviewing and “scoring” 16 graduate student modeling projects that were done either as part of their dissertation work or as class exercises. Model size, feedback complexity and behavioral complexity are generally comparable between the two samples (graduate students and H.S. students). The most striking difference is the higher scores on model development by the H.S. students! This is due primarily to the strong emphasis in the H.S. curriculum and in SyM Bowl on tying model parameters to real world data and on dimensional consistency. The other difference is the somewhat lower scores overall for the H.S. students regarding model testing. These differences are no doubt due primarily to the relative emphasis placed on these different aspects in the classroom. Clearly, there is room for improvement in both environments!

Conclusions

Due to the limited and highly exploratory nature of this work, any conclusions drawn must be, at best, very tentative in nature. The following conclusions are offered in this tentative spirit:

- High School students appear to generally do a good and often excellent job of tying their model parameters to real world data.
- Many of the best projects also carefully consider dimensional consistency and analyze the feedback structure of their model. Additional encouragement would be appropriate.
- The recent trend towards less complicated and better-tested models is good, and should be further encouraged by providing the students with plentiful examples.
- Better guidelines for model testing would appear to be called for, also in the form of well-documented examples.

The role played by SyM Bowl and its associated judging criteria and sample document is not as clear as one might hope. By closing the loop, it has no doubt had an impact. Most students follow the format suggested in the sample document **very** closely. When the criteria

“endogenous creation of the behavior of interest” was added, the number of table-driven models declined. When the emphasis on the clarity of the model diagram was increased, the number of “spaghetti” diagrams decreased. The students and their teachers **are** paying attention.

What are the strengths and weaknesses of the judging criteria and sample document?

- ❑ The criteria appear to be converging towards a very coherent, if cryptic, set of standards.
- ❑ The sample document on the other hand appears to be woefully inadequate.
 - While the format it offers, along with its other suggestions, is valid, it appears to stifle student creativity when it is followed too closely.
 - Furthermore, this past year, the *process* used to build models was de-emphasized in the criteria without increasing its emphasis in the sample document, the result of which was less information about the process in the student papers and poster sessions.

How might the judging criteria and sample document be improved?

- ❑ Add *learning about the subject of interest* to criteria 1 and criteria 7.
- ❑ Create clearly written paragraphs describing each criteria in greater detail.
- ❑ Instead of providing a single “dummy” sample document, provide dozens of example documents from different fields in order to illustrate the wide variety of situations and associated modeling processes in which SD modeling flourishes.
 - provided via the web
 - each exemplifying superior modeling and model documentation

These improvements are already “in the works,” and additional suggestions for improvement are being actively sought.

The SyM Bowl organizers, including the author, recognize the need to enhance the feedback loops shown in Figure 1 by strengthening communications between the organizers of the event and its participants, the H.S. students and their teachers. We began this process immediately after the event this year, and we plan to continue to meet monthly.

References

Gallaher, Wakeland, Fisher (2000) SyM Bowl 2000 Sample Paper. May be downloaded at:
<http://www.sysc.pdx.edu/download/papers/SB2000Sample.doc>

Wakeland (1998) "The Judging Process for SyM Bowl: a High School System Dynamics Modeling Competition," presented at the 1998 System Dynamics Conference, Quebec, Canada. May be viewed at:
<http://www.sysc.pdx.edu/download/papers/Sdconf98.htm>

Wakeland (2000) "Teaching Systems Science in High School Compared to Graduate School," to be presented at the 2000 I.I.S. World Congress, Toronto, Canada. May be viewed at:
<http://www.sysc.pdx.edu/download/papers/ISSS2000.htm>

Appendix A: SyM Bowl 2000 Judging Criteria

All criteria are equally weighted Items 1-8 are initially scored based on the written report.

"Good writing and good modeling go hand in hand." Jay Forrester 2/23/2000

1	The problem, purpose of model, and reference behavior pattern(s) are all clearly described.
2	The flow diagram is clearly laid out, and effectively shows the interactions between system elements. The names for stocks, flows, and converters are all descriptive and fitting.
3	Model feedback is appropriate, accurately identified, well-analyzed, and explained.
4	Model equations are simple (self-evident algebra) and clearly explained. Dimensions are correct and consistent. Assumptions are clearly explained and justified.
5	The model endogenously generates the behavior of interest. Exogenous inputs are identified, as appropriate.
6	The model is properly tested in that model behavior is compared with reference behavior patterns, test cases are run including extreme conditions, sensitivity analysis is conducted over a wide range of values, and dynamic behavior is thoroughly examined.
7	Model results and conclusions are clearly explained. Figures and Tables are concise and effective.
8	The report is well organized, grammatically correct, and rhetorically sound.
9	Poster clearly shows model purpose, model structure and feedback loops, key assumptions, behavior-over-time graphs, results and conclusions.
10	During the poster session, team members provide an informative overview, and are able to answer probing questions.

NOTE: Preliminary criteria scores from the written reports can be increased if students provide new information during the poster session.

Meaning of the Scores for a given criteria

4	Fully meets all aspects of the criteria without question
3	Generally meets all aspects of the criteria
2	Partially meets many aspects of the criteria
1	Partially addresses the criteria, but only in a very limited fashion
0	No evidence that this criteria was addressed or even considered

Standards:

Level	Total out of 40	Average Criteria Score	Minimum Criteria Score
Gold	35	3.5	3
Silver	30	3.0	2
Bronze	25	2.5	1