The Validation of Commercial System Dynamics Models

Geoff Coyle and David Exelby

HVR Consulting Services Ltd.

Selborne House, Mill Lane, Alton, Hampshire GU34 2QJ, United Kingdom.
Tel: 44 1420 87977
Fax: 44 1420 89819
david.exelby@hvr-csl.co.uk
geoff.coyle@btinternet.com

This paper discusses the need for formal criteria for the establishment of confidence in, or the 'validation' of, system dynamics models constructed for fee-paying clients as opposed to for academic research purposes. The meaning of 'validation' is first considered and the substantial differences between the consultancy and academic cases are discussed. That leads to a review of the system dynamics literature on tests of validity. Finally, there is a discussion of the process of consultancy in system dynamics.

This paper has been separately accepted for publication in System Dynamics Review in 1999.

The Consultancy Dilemma

The commercial consultant is paid by his clients to build models to help them to understand, and sometimes solve, problems. The decisions the client takes on the basis of the consultant's study may have financial and other ramifications which are so large that they are hard to compute. If the model is in some way 'wrong' the client may be led to erroneous conclusions and could justifiably be irritated, to say the least, that the consultant had apparently failed in his duty of care to ensure that the model is 'correct', or 'valid'. On the other hand, the consultant usually works to a tight budget, is under considerable time pressure and, significantly, *is dependent on the client for much of his information*. There are potential difficulties inherent in this situation and this paper discusses the problem of ensuring, as far as possible, the 'correctness' of the consultant's model. The discussion is first from a theoretical perspective and then more practically.

The Concept of Validity

A naïve client might believe that a model is true, or valid, because some test, such as a close fit to historical behaviour, has been satisfied. The corollary is that, prior to passing the test, the model was untrue, or invalid. The consultant therefore needs to make the client aware of the difference between truth and validity and the dictionaries give some guidance.[1] Valid means '(Of ... argument) sound, defensible, well-grounded. (Law) sound and sufficient, *executed with proper formalities*' [*Oxford*, emphasis added] or 'having premises and conclusion so related that whenever the former are true, the latter must also be true' [*Collins*, adding that 'in Logic, validity cannot be said to attach to a conclusion by itself']. defines true in several ways, the most useful for the purposes of this paper being '*conforming* with reality' [Emphasis added].

The consultant's model is, however, a *simplification* of reality which is intended to serve some useful purpose for the client. It cannot be said to be true or false, the most that can be said of it is that it is good enough for its purpose. Similarly, the Newtonian model of gravitation is useful for many purposes but has been found to be not good enough, or invalid, in some areas of modern physics. The phrase 'found to be not good enough' is a keystone of this paper and the concepts in it.

To the degree that a model passes tests that it is 'sound, defensible and well-grounded' it has that degree of validity and, hence, of being good enough for its purpose. If no tests were passed the model would be completely invalid and hence useless. A model might, however, pass many tests but fail one which is absolutely essential such as, in system dynamics, dimensional consistency. Such a model would be invalid as one would not know how much confidence could be placed in its outputs.

Overall, it might be better to refer to *the confidence that can be placed in the conclusions which can be drawn from a model's outputs*. However, to speak of its validity is more compact, provided care is continually exercised not to confuse validity with truth, both in one's own mind and, even more importantly, in that of the client.

Coyle (1977) summarises the point that validation is

> 'the process by which we establish sufficient confidence in a model to be prepared to use it for some particular purpose'.

This process has two aspects. One, validation, means ensuring that the model's structure and assumptions meet the purpose for which it is intended. The second, verification, means ensuring that its equations are technically correct, even to the extent of checking that + has not been used when - was required. This latter point is also called debugging and would be too obvious to merit mention had one not seen models in which precisely such mistakes had been made.

Validation subsumes verification. However, a verified model might be invalid if it correctly represented the wrong purpose. A valid model could not be unverified.

Risk and Defence: Some Differences Between Academic and Consultancy Work

The similarities and differences between professions are illuminating. A lawyer gives advice to his clients on matters which are important to them. A modelling consultant usually aims to help his clients arrive at decisions which may have far-reaching consequences and which often have financial ramifications. By contrast, an academic seeks to publish research results to a scholarly community.

These differences between academic and commercial system dynamics work can be expressed in the CATWOE mnemonic.[2]

	Commercial Consultancy	**Academic Research**
C Customer	Fee-paying client.	Research/Scholarly Community.
A Actor	Modeller **and** Client.	Modeller.
T Transformation	A client's problem into a viable solution.	A gap in knowledge into the basis for further development.
W Weltanschauung	Competition with other firms - Survival	Contribute to a body of scholarship
O Owner	Client.	Modeller.
E Environment	Competitive.	Fraternal.

The difference between the Owners has significant implications for validation. For the consultant, the client has selected and owns the problem and is usually much the more knowledgeable about the problem domain. This means that the client is the ultimate arbiter of validity *as he has a personal interest at stake*. Because of that, the client and the consultant jointly validate the model, the consultant verifying it. The academic, on the other hand, has chosen the problem himself and must verify and validate his own model.

In both cases, however, their work is open to inspection or review. Just as a lawyer who had not exercised proper diligence on behalf of a client could be in very serious trouble, a modelling consultant whose work was found to be not good enough could face serious penalties from an aggrieved client. Even when a consultant simply develops a model which is given to the client to use for himself the consultant's duty of care may increase as the client, if he is not expert in system dynamics, may draw incorrect conclusions and blame the consultant for an incorrect model. The consultant's risks are obvious and extreme; the academic's risk is of professional criticism if the work is found to be inadequate. Of course, if the inadequacies are grave, his or her professional reputation could be seriously damaged.

In consultancy, moreover, the risks are, in a sense, bilateral. The client must, at some point, have sufficient belief in the consultant's work to take action based on it, or he must write off the cost of the project. The consultant must, at some point, believe that his work is good enough to be presented to the client.

To account for validation and risk we may revise the acronym from CATWOE to CATWOVER:

	Commercial Consultancy	Academic Research
C Customer	Fee-paying client.	Research/Scholarly Community.
A Actor	Modeller **and** Client.	Modeller.
T Transformation	A client's problem into a viable solution.	A gap in knowledge into the basis for further development.
W Weltanschauung	Competition with other firms - Survival	Contribute to a body of scholarship
O Owner	Client.	The academic community.
V *Validation*	*Verification by consultant, Validation by consultant and client.*	*Verification and validation by researcher and possibly referees.*
E Environment	Competitive.	Fraternal.
R *Risk*	*For the consultant: financial, with possibility of litigation.* *For the client: potentially far-reaching if a decision is made which proves to be wrong as a result of an incorrect model.*	*Academic standing and reputation with peers.*

Given these differences, we propose a revised definition of 'validation' *for consultants* as:

the process of building up one's own confidence that a model is sufficiently good that one can risk using it as the basis for practical advice to a fee-paying client.

There is a corollary to this:

the process of building up the client's confidence that the model is good enough to be believed.

The second definition depends on the first but there are very clear behavioural implications in the second case. Their management forms part of the consultant's inter-personal skills and this paper will concentrate of the technical aspects of system dynamics practice which bear on the first of the definitions.

The essence of the matter is that one's work might be reviewed and, perhaps, criticised but, in some professions, there is a defence mechanism against accusations that the work was not good enough. Fortunately (for him) a surgeon whose patient dies on the operating table is not automatically sued by the relatives or struck off by the authorities. He merely has to show that he acted according to accepted medical practice. Thus an accusation that work is not good enough can, in principle, be refuted by showing that it was done in accordance with accepted standards. It is the purpose of this paper to develop a basis for accepted standards in system dynamics consultancy. First, it is necessary to consider review processes.

Review Processes

Review processes can take many forms.

In medicine, hospitals hold case conferences to allow advice to be given and advances in treatment to be made. Sometimes a doctor's standard of practice might be criticised. Legal practices hold similar reviews.

Academic research is reviewed, or quality controlled, by journal referees. They base their judgement on the clarity of the writing, whether due recognition has been given to previous work and whether the conclusions seem to follow from the apparent assumptions. In system dynamics they usually have no opportunity to investigate the model's equations; the most they are likely to see is a stock/flow diagram or an influence diagram so they have to take the graphs of results on trust. A key point is that the referees have no personal interest at stake.

Review of consultancy models is less formal. Firms which work for government are required to have quality assurance programmes. In the UK, firms submit themselves to inspection for a quality award such as ISO9000 and must then have a quality assurance system. They will also have a quality manager who, in practice, may well not have time to scrutinise the work in any detail and has to take much on trust. In any case, ISO9000 does not specify what a quality manager would have to do in, say, a confectionery manufacturer as opposed to a system dynamics consultancy.

The client has a personal interest at stake but is unlikely to have the technical resources to check the model and has to take much on trust. However, cases have been known where a client has hired another expert to scrutinise the first consultant's work. We have been involved in several such assignments.

It is therefore clear that a quality-assured consultancy requires an internal process of model scrutiny which is as rigorous and impersonal as a medical case review. Such a process would have six benefits:

> it would ensure as far as possible that there were no mistakes, such as structure unsuitable for the model's purpose or incorrect code, hidden in the model waiting for discovery and subsequent complaint by an alert client or independent scrutineer;
>
> it would provide modellers and their supervisors with a handbook of good practice;
>
> the knowledge that good practice was being followed should lessen the load on supervisors and quality managers;
>
> it would ensure greater consistency of work standards across the firm;

it would act as a marketing tool in that the consultancy could demonstrably claim to follow a theoretically sound and clearly defined process rather than making general assertions about quality assurance;

it would give clients a basis for evaluating bids from competing consultants, at least as far as their respective quality assurance procedures is concerned.

In short, the process would specify the proper formalities mentioned in the dictionary and would be a defined series of tests that a model was sound, defensible and well grounded. The more tests had been passed, the more confidence could be placed in the model's validity. Some of the tests would be mandatory in the sense that every model would have to pass them. Others might be elective in that they would not necessarily apply to all models. The modeller would, however, have to give the reasons why Test X did not apply to Model Y. The test record for a given model would form a useful archive of that model and be a source of training and experience building for new recruits.

To develop the process it is first necessary to establish its principles by reviewing the system dynamics literature's discussion of validation, in chronological order of publication.

The System Dynamics Literature on Validation

Forrester (1961) lays great emphasis on validation, posing the question of whether there are objective and uncontroversial tests that a model is properly constituted. Those pages contain a wealth of wisdom which can be summarised only briefly here.

His key point is that the defence of a model must rest on the defence of its details. The behaviour of graphs of selected variables is no guarantee of validity as 'an endless variety of invalid components [equation forms] can exist to give the same apparent system behaviour'. He insists that the equations must be dimensionally consistent and that all the constants in a model must be clearly defined and their dimensions must be stated.

Forrester goes on to state that the behaviour of a model should be carefully checked, as 'serious model defects will usually expose themselves through some failure of the model to perform *as would be expected of the real system*' [emphasis added]. He is uncompromising that 'improvements must be made only if they represent the real system, not because they fix the problem'.

Throughout, Forrester is at pains to emphasise that there are no absolutely objective tests that a model is acceptable. At the heart of the testing process is an inherent assumption that the tests being applied are necessary and relevant to the purpose of a particular model and, in general, to the practice of system dynamics. To that, we would add that tests can only be termed 'objective' if they do not depend on the emotions and biases of the tester and they are widely accepted by the community of practitioners. In other words, that they constitute the proper formalities, as defined in the dictionary.

Coyle (1977) considers validation from the similarly pragmatic point of view that ‘there are almost endless opportunities for making mistakes in any kind of model building’. He states the principles of dimensional analysis and illustrates their use to detect errors in models

He anticipates later work by others in a general discussion of validation, giving a list of criteria. Briefly these are that the model’s boundary should be appropriate to its purpose; that there should be no gross errors of behaviour (which might, for example, arise from confusing decimal values and percentages or, particularly, from dimensional errors); that the equations and parameters should correspond to the real system; and that the model should reproduce the system behaviour in various conditions.

Finally, he remarks that ‘we feel that the best test of confidence in a model is the knowledge that it has been carefully built up in conjunction with the management’. This close proximity to the demands and knowledge of a client with a personal interest at stake is the essence of the difference between the consultancy and academic facets of system dynamics.

Forrester and Senge (1980) propose objective and uncontroversial tests of validity. They stress that a model is built for a purpose and its validity is fundamentally determined by the extent to which it fulfils that purpose. They lay much emphasis on the boundary of the model, recognising that a model is a simplification and that the boundary between what has been included and what has not is a significant determinant of the model’s validity. They then propose a series of tests of the model’s structure, behaviour and policy implications.

Lyneis (1980) explains standard test inputs such as STEP, NOISE etc.

The standpoint of Richardson and Pugh (1981) is that ‘a system dynamics model addresses a problem, not a system and is designed to answer a reasonably well-defined set of questions’. They pose the tests of whether a model is ‘suitable for its purpose and … consistent with the slice of reality which it tries to capture’. They make the profound point that these questions are best posed in the negative; ‘in what ways is the model **un**suitable for its purpose’ etc. They restate Forrester and Senge, adding the criterion of whether the structure of the model is suitable for its audience.

Sterman (1984) proposed statistical techniques for assessing the quality of fit between a system dynamics model and historical data. Wolstenholme (1990) and Mohapatra *et al* (1994) restate the Forrester and Senge tests.

Lane (1995), summarised and simplified by Pidd (1996), discusses validation in the context of different modes of use of system dynamics models.

Barlas (1996) develops the idea that system dynamics validation is rooted in a relativist rather than an absolutist point of view, though that does not exclude the use of formal tests. He supports Forrester’s view that the defence of a model rests on the defence of its details. Barlas repeats the Forrester and Senge criteria, adding the so-called Turing test in which managers are shown randomly sorted outputs from a model and historical behaviour and are asked if they can tell the difference. Such a procedure would, however, probably require double-blind testing and a large sample and might lead people to rest the defence of the model on its output, not on its details. Barlas discusses the limitations of the application of statistical tests to the outputs of a system dynamics model in comparison with real-world data and proposes formulae for such tests when it is appropriate to use them.

Coyle (1996) discusses validation throughout the book, validity of a model being defined as 'well-suited to its purpose and soundly constructed'. His key philosophy is that 'the model should do the same things as the real system *and for the same reasons*'. He lays absolute emphasis on dimensional consistency, the rules for which are stated. He emphasises the need for correspondence between the model and the problem, though stressing technical aspects such as DT invariance and mass balance. He adds that all constants must be properly defined and their dimensions must be stated. (See also Forrester, 1961)

Coyle emphasises the role of the influence diagram as a tool for thinking about the problem. He proposes the 'cone' of diagrams shown in Figure 1. In view of its relevance to validation, this must be briefly discussed.

In Figure 1, each ellipse represents an influence diagram drawn at an appropriate level of detail, each being conceptually consistent with its lower and higher neighbours. The diagonal ellipse suggests a diagram which shows some parts of the system in detail and others at a broader level. The level 4 diagram will usually be very much larger than level 3. In some cases, it is suggested that a model can be entirely qualitative, consisting only of an influence diagram (Coyle, 1998; Wolstenholme, 1999).

It is argued that a study will normally commence at level 2 and will progress to the lower levels as the study proceeds, though often iterating to higher levels. At the conclusion of the study, its main results should be expressible at level 1. He proposes 10 tests of what constitutes a 'good' diagram.

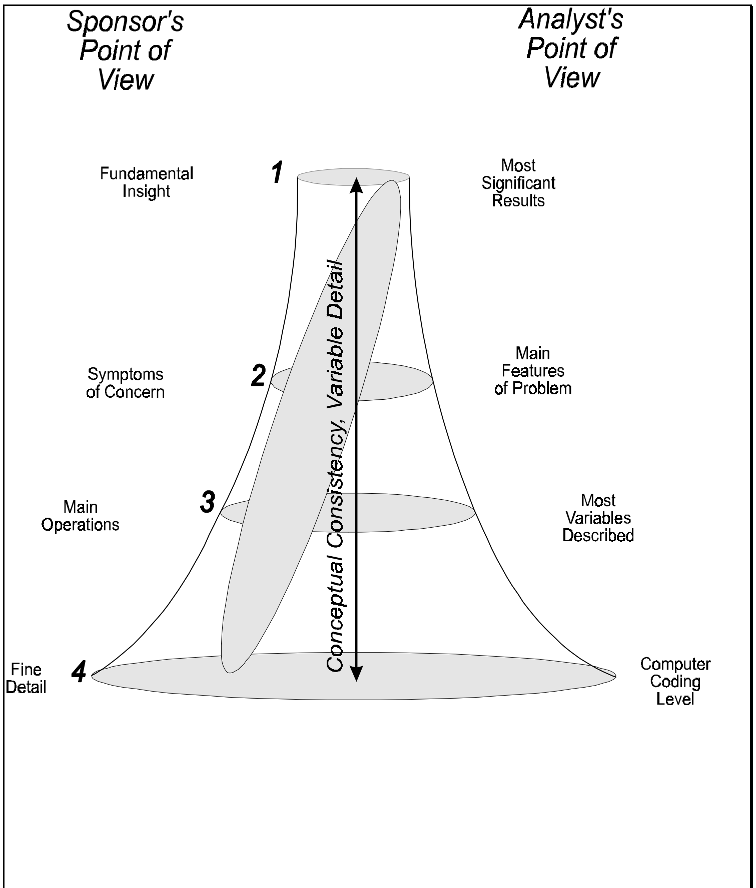

Figure 1 The 'Cone' of Influence Diagrams

The cone's relevance to validation lies in the distinction it makes between the sponsor and the analyst. In academic research, the analyst is his own sponsor and must learn enough about the problem domain to be able to stipulate the symptoms, operations and detail on the left hand side of Figure 1. In consultancy, the client is both the sponsor and the domain expert, not the consultant. Thus, the academic validates his own work, being sufficiently expert in the technique to verify its technicalities. The consultant, on the other hand, must work closely with the client to validate the model though, as the client is not expert in the technique, the consultant must verify its technicalities. The client requires assurance that this has been done so the consultant must use rigorous formal model tests, the record of which can be shown to the client if necessary as evidence that the work has been executed with proper formalities.

Coyle also proposes general criteria for assessing the quality of a model as a whole. These build on the criteria proposed in Coyle (1977) and continued in Forrester and Senge (1980) though adding new ideas. He proposes 15 general headings under which a model's validity might be assessed.

Finally, Richardson (1996), in a wide-ranging discussion of problems in system dynamics, calls for effort to develop techniques and procedures for validation. He refers to the Forrester and Senge tests but cites evidence that many of the tests are little used by system dynamics practitioners (Scholl, 1992). In his view validation has to be seen in the context of the use of a model: a model used for testimony in court might, he argues, require a level of validation different from one used by a group of executives. One might differ from that view and suggest that the importance of validation resides in the significance of the issue being addressed and the consequences of error. However, Richardson calls for more work on validation and it is hoped that this paper has made a contribution to that task.

Summary of the Literature

This review of the academic literature on the validation of system dynamics models suggests two things. One is that the attitude of authors to this subject is by no means consistent. Some treat it in detail, others more cursorily.

Of those who do consider validation in detail (Forrester and Senge (1980) and Coyle (1977, 1996)) there is a broad measure of agreement, the main heads being:

> there is no such thing as absolute validity, only a degree of confidence which becomes greater as more and more tests are passed;
>
> some tests are so significant that they can be regarded as mandatory, dimensional consistency being a *sine qua non*;
>
> the validity of a model can only be assessed in relation to its purpose; it must be capable of answering the questions its sponsor wishes to ask;
>
> the boundary between what has been included and what has been omitted is crucial;
>
> the boundary must be small enough to be tractable yet large enough to encompass the solution to the problem;[3]
>
> validation by defence of the detail is essential because there are limitless opportunities for making mistakes, especially as there is often a paucity of theory to guide the analysis;[4]
>
> the behaviour of the model must be like that of the real system and changes to the model's structure, parameters and policies must produce changed behaviour which is plausible and explicable; this must still be the case even when the changes are extreme.

The above is a summary of the theory. The problem is that the criteria are broad and general and, while they are essential as an attitude of mind, they do not go far enough to support the commercial modeller. Such a person requires more detailed and specific tests which can be shown to a client or supervisor as evidence that a model is relevant to the problem, soundly conceived, well-constructed and executed with proper formalities.

Verification and Validation in Consultancy

To understand the concept, it is useful to describe the process of consultancy, as shown in Figure 2, the notation of which must first be explained. Three of the four cloud shapes, Real World System, Real World Behaviour and Model Parameters represent areas from which information must be extracted. The fourth, Clients' Mental Models, represents the understanding of the processes in the system which the clients have, probably unconsciously, formulated from their accumulated experience of the real system. Note that clients is expressed in the plural as, usually, no one person in the client organisation fully understands the problem. In any case, there may be complementary or competing mental models elucidation of which should add richness to the analyst's understanding.

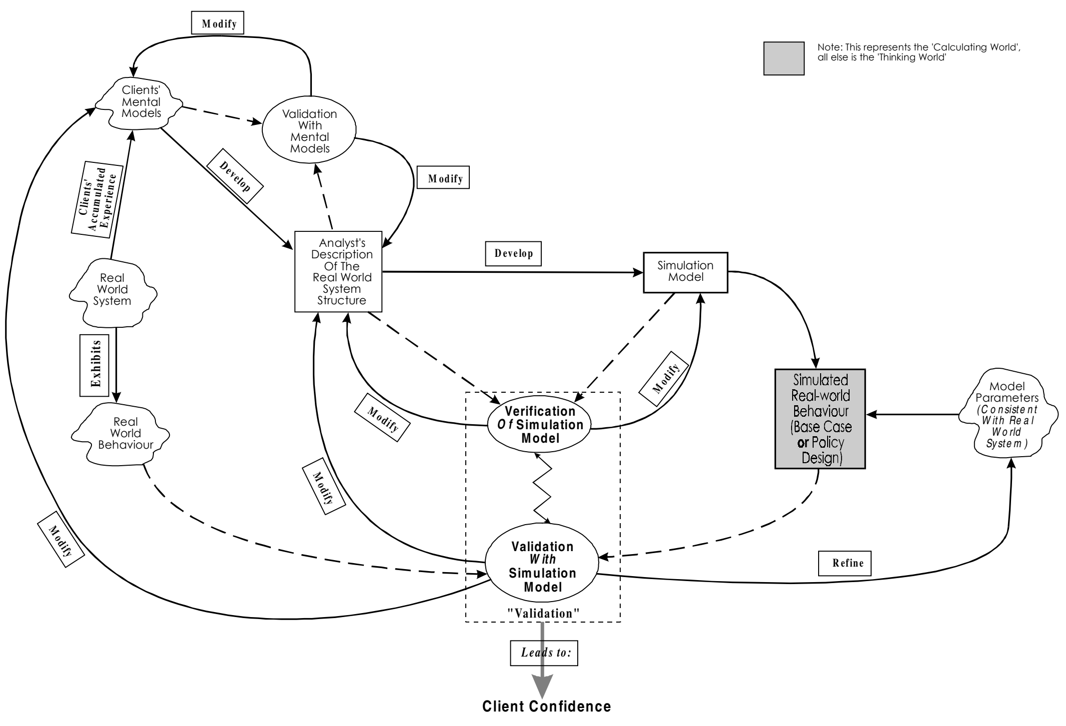

Figure 2 Validation, Verification and Modelling

The most difficult task of consultancy is to understand the problem in hand by elucidating the mental models and transforming them into influence diagrams. This task is supported to some extent by the process of parsing notes of interviews, company documentation and so on (Coyle, 1996). This exploits Forrester's insight that there is very much more useful information in company records and peoples' minds than in statistical data. The consultant must, however, get inside his clients' minds; the academic needs only to use his own brain.

The three rectangles with arrows coming in or going out are the stages at which the analyst performs technical work using defined procedures (such as the three methods of building influence diagrams described in Coyle, 1996).

The solid lines show sequences of work and are labelled to indicate broadly what happens during that step. The three ellipses are the places at which checks are made that a previous stage of the modelling has been correctly carried forward to the next. The stages compared are denoted by dashed lines.

The three comparisons are:

Validation with Mental Models. Has the problem been well-understood, has a suitable boundary been chosen and can the problem be turned from the clients' imprecise mental models, in the plural, into a rigorously defined Influence Diagram?

Verification ***of*** *the Simulation Model*. Has the influence diagram been correctly translated into a simulation model? Does that model conform with the theory of dynamic systems and does it satisfy the criteria used within a particular consulting firm? This is done by the technical analysts.

Validation ***with*** *the Simulation Model.* At this stage the model's behaviour is compared with the real world. This is often done by the project supervisor (rather than the actual modeller), *and* the client, recalling that the client is the domain expert. It is important to stress that this is done not only for the base case behaviour, when one is often, to some extent, comparing the model with historical behaviour. It is also done when the effects of policy simulations or other changes to the system or its structure are evaluated for reasonableness, even though there is, by definition, no history against which to compare.

Both of the last two steps are subsumed into the overall heading of validation which, if achieved, should lead to the consultant and the client having confidence that the conclusions drawn from the model can be trusted.

The processes shown in Figure 2 translate into four broad stages of modelling work

In the *Thinking World* the model is conceptualised and the problem understood. The output of this stage is a well-understood problem, a clear diagram of it and a planned approach to modelling it.

The *Transition from Thinking to Calculation* has two phases. In the first, the understanding of the problem, reflected in the diagram, is translated into equations in one's head or on paper, *without using the computer*. In the second phase, the equations are loaded into the software and the initial model is tested and checked.[5] The output from this should be a model which is accepted as being able to do the same things as the real system and for the same reasons with respect to some base case conditions.

In the *Calculating World* the objective is policy design. Experiments are performed on the model, one of which is to subject it to extreme conditions. The output of this stage is the basis for recommendations to the client. In practice, errors may be found in this stage which should have been detected and corrected in the previous stage.

In the *Communication World* the results are expressed in a format which can be presented to the client. The result should be a satisfied customer who thinks the work is good enough to be believed and who is prepared to pay the bill. It is, of course, very unlikely that the model itself will bring about change. The very fact of a study having taken place is likely also to have its effect on the client organisation and its policies; a strong case of Heisenberg's uncertainty principle.

In practice, each the four stages is likely to be revisited during a project. Generally, the validity of a model should steadily improve until it reaches a satisfactory level. This is akin to the concept of iterative conceptualisation and testing discussed by Randers (1980). The second and third stages do not exist in a purely qualitative study. The third stage, policy design, might not exist in a project to develop a model to be delivered to the client for his own use. In such a project the fourth stage, communication, would probably be more a matter of preparing documentation and training for the client so that he can use the model with confidence and in full knowledge of what it *cannot* be expected to do, as well as what it can.

The reason why academic research differs from this process is that the researcher is unlikely to be as expert in the problem domain as is a consultant's client. In some cases, one has to hypothesise a structure for the model. As Richardson and Pugh (1981, page 55) express it, 'The dynamic hypothesis in a system dynamics study is a statement of system structure that appears to have the potential to generate the problem behaviour'. Such an approach therefore starts with the Real World Behaviour in Figure 2 and then moves to the diagrammatic model of the chosen structure. This might seem to make it difficult to use the real world behaviour as an aspect of validation of the structure as *the structure was assumed from the behaviour*. Other structures might, just as reasonably, have been selected so, in short, the hypothesised structure may be a good one but there is no way of knowing whether there is a better one still.

In fact, the academic modeller will also use other information, or assumptions, to support the structural hypothesis and is very likely to do so in developing rate equations. Nevertheless, the academic modeller has selected the problem himself and is much more his own master than is the consultant. The latter must ensure that, as far as possible the *clients'* knowledge is represented in the model. Figure 2 therefore emphasises that the consultant is more closely constrained to use the real world behaviour as a test of what has been discovered from its originating system, not from some independent hypothesis.

Figure 2 and the four phases of modelling are still in general terms and to convert it to operational reality in a consultancy firm will require a set of applicable criteria. Although rooted in the academic theory, such tests would be orientated towards the detailed justification that, within its boundary, the model *does the same things as the real system and for the same reasons*. In day-to-day use on consultancy projects they would form the basis of formal documentation of a model.

A given consulting firm will be likely to have its own procedure, the details of which will form part of its stock in trade and will, therefore, be commercially valuable and confidential. In the case of the company with which the second author is associated, the validation and verification processes are supported by a total of 75 separate tests under 44 headings in the 4 stages with detailed explanatory notes. In addition, there are 10 tests of a good influence diagram and 15 tests of the overall quality of a model, both of which were mentioned earlier.

Summary and Conclusion

The paper has discussed the question of validation in system dynamics from a theoretical point of view and has related that to the problems of commercial consultancy. It was argued that consultants are in a different position from academics and require more formal validation procedures.

Of course, quality assurance is a continuing and inherent part of the modelling process. Whatever set of detailed criteria a given consultancy firm might use has to be viewed as a manual of good modelling practice which is part of the individual consultant's normal working technique for models on which she is working. The criteria lose their purpose of good practice if they become *ex post facto* tests on a model which has been built.

In cases where the consultancy has been asked by another firm's client to review that firm's work, and such instances are by no means uncommon, the tests would be the set of criteria to be used in the evaluation. The overall validity of the model would be assessed in the light of the tests it had passed and those it might have failed.

References

(Note: Books originally published by MIT Press and Addison-Wesley have been re-issued by Productivity Press)

Barlas Y (1996), 'Formal aspects of model validity and validation in system dynamics models', *System Dynamics Review*, Vol. 12, No. 3 (Fall), pp. 183-210.

Coyle R G (1977), *Management System Dynamics*, Chichester, John Wiley and Sons.

Coyle R G (1996), *System Dynamics Modelling: A Practical Approach*, London, Chapman and Hall.

Coyle R G (1998), 'The practice of system dynamics: milestones, lessons and ideas from 30 years experience', *System Dynamics Review*, 14(4) 343-365.

Forrester J W (1961), *Industrial Dynamics*, Cambridge, MIT Press.

Forrester J W and P Senge (1980), 'Tests for building confidence in System Dynamics models', in Legasto A, J W Forrester and J M Lyneis (editors), 'System Dynamics', *TIMS Studies in the Management Sciences*, 14, pp 209-228.

Lane D C (1995), 'The folding star; a comparative reframing and extension of validity concepts in system dynamics', *Proceedings of the 1995 International System Dynamics Conference*, Tokyo

Lyneis J M (1980), *Corporate Planning and Policy Design*, Cambridge, MIT Press.

Mohapatra P K, P Mandal and M C Bora (1994), *Introduction to System Dynamics Modelling*, Hyderabad, Universities Press of India.

Pidd M (1996), T*ools for thinking - modelling in management science'*, Chichester, John Wiley and Sons.

Randers J (1980), 'Guidelines for model conceptualisation', in *Elements of the System Dynamics Method'*, (ed J Randers), MIT Press.

Richardson G and A L Pugh (1981), *Introduction to System Dynamics Modelling with DYNAMO*, Cambridge, MIT Press.

Richardson G (1996), 'Problems for the future of system dynamics', *System Dynamics Review*, Vol. 12, No. 2, pp. 141-1157.

Scholl G (1992), 'Benchmarking the system dynamics community', *System Dynamics Review*, 8(3), 263-266.

Sterman J (1984), 'Appropriate Summary Statistics for Evaluating the Historic Fit of System Dynamics Models', *Dynamica*, Vol. 10, Part II, Winter 1984.

Wolstenholme E F (1990), *System Enquiry. A System Dynamics Approach*, Chichester, John Wiley and Sons.

Wolstenholme E F, (1999), 'Qualitative *versus* Quantitative Modelling: the Evolving Balance', special issue on system dynamics of *Journal of the Operational Research Society* to appear.

---

[1] The sources are *The Shorter Oxford Dictionary* and *The Collins English Dictionary*.

[2] The term derives from Checkland P, *Systems Thinking, Systems Practice*, John Wiley, 1981. One can do no better than quote directly: 'C represent the customers who are the victims or beneficiaries of a transformation. A is the actors who do the transformation, T is the process of transforming an input into an output, W is the 'Weltanschauung' or world-view which makes the transformation meaningful in context, O are the owners who could stop the transformation, and E are environmental constraints or elements outside the system which it takes as given.'

[3] There are concerns that the modern tendency to build large models has the effect that tractability and solvability are buried in illusory detail, apart from which there is an exponential increase in difficulty of debugging as the model grows larger.

[4] Coyle (1996, Chapter 12) points out that the system dynamics practitioner is always faced with a blank piece of paper at the outset of a study, unlike the linear programmer who knows in advance that the model must consist of a set of linear constraints of certain types and a linear objective function.

[5] In practice, the nature of the software and the experience of the modeller often means that phase 1 is also carried out on the computer. Ideally, only experienced modellers should do this. The essential thing, even for the experienced, is to maintain in one's mind the distinction between the two phases.