Evaluating long-term impact of qualitative system dynamics workshops on participant mental models

Rodney J Scott
Robert Y Cavana
Donald Cameron

The University of Queensland
Brisbane QLD 4072 Australia
rodney.scott@gmail.com

Several studies that describe group model building interventions assert that these interventions can produce new insight. No previous study has examined whether these new insights are enduring. This paper reports on evaluation of four qualitative system dynamics interventions conducted 12 months after the interventions. The interventions each consisted of a three hour workshop followed by a two-hour workshop one week later, to plan strategy implementation initiatives in a government department. A change of circumstances meant that the workshop conclusions were not implemented. Statistical comparison of work samples from immediately before and after the workshop, and twelve months later, suggest that participants views on the workshop topic changed through the workshop process, and that these changes were enduring even in the absence of reinforcing activities. Mental model change is proposed as an explanatory mechanism.

Key words: Qualitative system dynamics, group model building, evaluation, long-term, insight.

INTRODUCTION

Doyle and Ford (1998) propose that a key challenge for the group model building community is to establish the stability of any changes in mental models brought about by brief intervention. This study reports on a group model building project, where the conclusions were not implemented, and consequently presents a relatively unmodified environment to observe the impact of a brief intervention on participants thinking.

Group model building literature remains unclear on what these methods achieve, and why (Rouwette *et al.*, 2011). Several authors have explored hypotheses for how group model building encourages mental model *change*, without knowing whether these changes are enduring (Richardson *et al.,* 1994, Vennix *et al.,* 1996, Maani and Marahaj, 2003, Rouwette *et al.,* 2011).

Previous studies establish that brief interventions can produce insight and mental model alignment among participants (Rouwette *et al.,* 2002, Scott *et al.,* 2012). Evidence of lasting insight and mental model alignment is a further step in understanding the impacts of group model building methods, and may aid in both the planning of future projects and the refinement of these methods.

This paper consists of five sections after this introduction. First, there is a brief re-cap of group model building evaluation literature. Second, there is description of the case study context and post-workshop results (previously reported in Scott *et al.,* 2012). The third section describes the evaluation methodology for understanding long-term impacts. The fourth section reports on the results observed through these

evaluations. Finally, there is a discussion of mental model change as a possible explanation for the results observed, and opportunities for further research.

GROUP MODEL BUILDING EVALUATION

Proposed interventions identified through system dynamics modelling are not automatically adopted by an organisation (Rouwette and Vennix, 2006). This may be due to a lack of understanding of prevailing politics (Greenberger *et al.,* 1976), or a lack of ownership by the client (Stenberg, 1980). As a result, some practitioners experimented with involving client groups in the modelling process (Richardson *et al.,* 1994). These approaches are now commonly referred to a "group model building" (Rouwette *et al.,* 2002) or "participatory modelling" (Rouwette and Vennix, 2006).

Most studies used anecdotal or descriptive evidence in evaluating group model building – only a small number attempted quantitative assessment (Rouwette *et al.,* 2002). Most commonly, these take the form of questionnaires conducted immediately after the workshop (Rouwette *et al.,* 2002). Other authors have explored behaviours observed during the workshops themselves (McCardle-Keurentjes *et al.,* 2009, Franco and Rouwette, 2011). There has been little study on whether these interventions are associated with long-term systemic impacts (McCartt and Rohrbaugh, 1995, Zagonel and Rohrbaugh, 2007) or participant learning (Cavaleri and Sterman, 1997, Thompson, 2009).

Based on the review of Andersen *et al.* (1997) of the existing group model building literature, Rouwette *et al.* (2002) identified a number of outcomes that were generally described as beneficial, including "learning or mental model refinement" and "mental model alignment". Mental models are explanatory schema that individuals use to explain the world and their interaction with it (Johnson-Laird, 1995), and that are relatively enduring (Doyle and Ford, 1998). In order to demonstrate that views changed through group model building can be equated to mental model change, it is necessary to establish that these changes can endure.

There have been some evaluations of long-term impacts in other fields that use workshop interventions (Decker *et al.,* 1988, Steinert *et al.,* 1993, Rust, 1998), involving delays of between 6 and 30 months. These generally report that positive outcomes occurred after an intervention, but there have been few attempts to demonstrate that the outcomes can be causally attributed to the intervention (Shadish *et al.,* 2001).

While there is good evidence that group model building is effective in producing improved communication quality, insight, consensus and commitment to conclusions immediately after the workshops (Rouwette, 2012), evaluation of long-term impacts is important for understanding the value of group model building interventions (Doyle and Ford, 1998).

CASE STUDY

Group model building interventions were applied to four multi-disciplinary groups selected from across a large government department in New Zealand, to plan the implementation of a recently developed long-term strategy.

Each group was tasked with identifying actions to implement one aspect of the strategy. Each group participated in a 3-hour qualitative system dynamics workshop (reported in detail in Scott *et al.,* 2012), and then a 2-hour workshop one week later. The workshops were both facilitated by an organisation employee.

The first workshop followed the methodology described by Maani and Cavana (2007), and consisted of the following components:

- Define the problem or situation
- Identify key variables
- Identify behaviour over time of main variables
- Construct causal loop diagrams
- Identify key leverage points for intervention by the organisation

The second workshop consisted of reflecting on and refining the conclusions from the first workshop.

Several sources of data were collected during this process. Before the first workshop, participants were asked to respond in writing to the question "What are the most important actions for (the organisation) to achieve (strategic goal)?" with four spaces (boxes) for answers. Following the first workshop, participants were asked this same question again, and also given a questionnaire that evaluated participants' views on the contribution of the workshop to increased communication quality, insight, consensus and commitment to conclusions. Session notes were also taken by the facilitator. The results of these evaluations are described in Scott *et al.* (2012).

The original research design called for ongoing interviews to determine whether the actions identified through the intervention were implemented. Immediately after the intervention, the case study organisation underwent a major merger and restructure. This resulted in the organisation not proceeding with strategy implementation in the way that had been intended. The magnitude of change was such that it would no longer be possible or applicable to evaluate the implementation of actions chosen. The results of the workshop were not implemented and were not formally discussed again.

METHODOLOGY

One year after the original intervention (between 51 and 53 weeks for each participant), the original participants from the four groups were asked to be part of semi-structured interviews. The time frame of one year was selected as representative of other studies that evaluate long-term impacts of workshops, that varied from 6 to 30 months (from other disciplines - Decker *et al.,* 1988, Steinert *et al.,* 1993, Rust, 1998). Of the 40 participants who had completed the original evaluation in full, 31 were still working for the organisation, including one who was unavailable due to parental leave. All 30 employees present in the workplace were interviewed individually. Participants who were likely to encounter one another in their daily activities were interviewed in quick succession to reduce the likelihood that they would gain prior knowledge of the interview questions.

The interviews covered five themes; the first two concerned the workshops themselves:

1. Participants' recollections of the workshops
2. Participants' beliefs about whether (and how) the workshops changed their views

The remaining three themes concerned the organisational strategy. Each group had participated in workshops concerning one of the four strategic objectives in the organisational strategy. Exploring how participants experienced the other three objectives (that weren't discussed during the intervention) provides some insight into the counterfactual – how participants may have changed their views without the intervention.

3. Extent of participants' knowledge of the organisation strategy
4. Extent and nature of impact of the organisational strategy on participants' work
5. Extent and nature of changes over time to participants' beliefs on the strategy.

The interviews also included three written questions. The first was a repeat of the question asked immediately before and after the initial workshops: "*What are the most important actions for* (the organisation) *to achieve* (strategic objective)?", again with four spaces (boxes) for answers. This is subsequently referred to as the "*action list*" method.

The second written question followed the format: "*Consider the following statements. Which of these do you think are the most important for* (the organisation) *to achieve* (objective)? *Please rank them from 1 to* (n)." Here, the participants' pre- and post- workshop answers were supplied, in random order (where "n" is the total number of unique answers provided by that participant in their pre- and post-workshop responses). Actions identified both before and after the workshop were removed. Participants were not told that they were the authors of these questions, nor which came from the pre- or post- workshop answers. Answers were ranked, rather than selected (eg: "select the 3 most important"), as participants sometimes listed a different number of actions before and after the workshop and relative ranking allowed easier quantitative comparison. This is subsequently referred to as the "*rated preference*" method.

The final written question followed the same format, however the objective was not the one that had been the subject of their workshop, and the statements were randomly chosen from the workshop that had considered that objective. This acts as a control tool (see results and discussion), and is subsequently referred to as the "*control ratings*" method.

RESULTS

Interview results

The interviews were transcribed and coded using content analysis (Holsti, 1969, Cavana *et al.*, 2001), consisting of five steps:
- Read through the responses
- Code themes as they emerge (open coding – Strauss and Corbin, 1990)
- Check material coded to each theme for consistency and clarity (axial coding – Neuman, 1997)
- Identify rules for inclusion/exclusion from the themes
- Recode responses based on rules for inclusion/exclusion (selective coding – Neuman, 1997)

Where several interviewees describe the same concept, direct quotes have been selected that are most representative of multiple respondents. A single coder coded all results.

The interviews revealed that most (17 of 30) had little recollection of the intervention, and could not describe the workshop in detail. Two described features that had not been part of the workshop (an explanation of the strategy and how it would be implemented, and a supplied definition of the strategic objectives and their scope) – it is possible that these were featured in other unrelated meetings in which the participants had taken part.

Among those who could describe the meeting in detail, follow-up questions revealed that most often (10 of 13) these participants had subsequently participated in other workshops that followed the same format. It was not always clear whether they were describing the intervention being evaluated or another similar workshop.

Few (6 of 30) recalled changing their views because of the workshop, though a larger number (14 of 30) believed that other participants had changed their views to become more like those of the interviewee. Where participants could describe the meeting, several themes emerged. Common themes were that participants believed:

- that they were listened to and understood ("*It was good when people suggested arrows* (in the causual loop diagram), *they could see their idea being used.*");
- that all could contribute to the discussion ("*People didn't need to be experts in everything...they could still contribute.*");
- that seeing things represented visually was useful in building shared understanding ("*It helps* (the discussion) *when you're all looking at the same picture.*");
- that focussing on causal relationships was useful in clarifying the participants own thinking ("*Focussing on causes is really good, not just a bunch of statements.*"); and
- that intervention points represented a shared and agreed path forward *("It was easier to pick interventions using the board* (causal loop diagram*) than just arguing it out against each other.*").

Questions concerning the organisational strategy revealed that most participants had familiarity with each of the strategic objectives, though this took some prompting. Few (7 of 30) believed that the strategy had changed their work, though some of these (4 of 30) believed that the organisation had shifted its focus from sustainability outcomes to economic outcomes (this shift is not mentioned in the strategic objectives). Most participants did not discuss the strategy regularly (21 of 30).

Participants believed that their views of the strategic objectives and what the organisation should do to achieve them had not changed for three of the four objectives. For the fourth objective "Increase sustainable resource use", participants believed that their views had changed based on a different interpretation of the objective (initially some believed it referred to increased sustainability, but more recently they understood it to mean increased resource use).

"Action list" method

The same question had now been asked of participants three times – immediately before the intervention, immediately after the intervention, and one year later. The written responses for all three sets were coded using content analysis (see above, Cavana *et al.,* 2001). Those answers by participants that did not take part in the delayed evaluation were removed from the analysis (resulting in data that is slightly different than that reported in Scott *et al.,* 2012). Participants were significantly more (p<.01) likely to volunteer responses that were coded the same as their post-workshop evaluation than their pre-workshop evaluation (see Table 1), suggesting far greater retention of post-workshop responses. This finding is re-tested later using another evaluation method (see also Tables 3 and 4).

Table 1: Origin of delayed responses, "action list" method

	Group 1	Group 2	Group 3	Group 4	Total
From pre-	0.24	0.18	0.06	0.11	0.16

workshop evaluation					
From post-workshop evaluation	0.48	0.44	0.38	0.57	0.48
New	0.28	0.38	0.56	0.31	0.36

Fewer answers were volunteered in the post-workshop evaluation. In the earlier paper (Scott *et al.,* 2012), this result was attributed to the use of leverage points to focus attention on the "critical few" points of intervention. In the delayed evaluation, this number had rebounded, suggesting that this effect may be time-limited (see Table 2).

In the pre-workshop, post-workshop and delayed evaluation, some responses were volunteered by more than one participant. While the delayed evaluation revealed answers slightly less alike than the post-workshop evaluation, both were still significantly higher (p<.01) than the pre-workshop evaluation (see Table 2). This was also true as a percentage of total answers (see "agreement rate" – number of participants who offered each response, divided by the total number of responses, Table 2). This suggests alignment of mental models among participants persists until at least 12 months after the intervention.

Table 2: Frequency of each response

	Total number of responses	Number of participants who offered each response	Agreement rate	Number of respondents
Pre-workshop evaluation	98	1.84	0.0187	30
Post-workshop evaluation	85	4.25	0.0500	30
Delayed evaluation	116	4.14	0.0357	30

There was no significant variation (p<.01) by age, gender, education, tenure or rank. This contrasts with the post-workshop evaluation (reported in Scott *et al.,* 2012) that saw workshop elements rated higher by non-managers; power-levelling (Van Nistelrooj *et al.,* 2012) was identified as a possible mechanism. This suggests that while group model building helps non-managers to contribute, there was no significant difference between whether managers' and non-managers' views changed through the intervention and if these changes endured.
Rouwette, *et al.,* (2011) proposed that participants in group model building processes learned from other participants ("transfer"), and from novel ideas emerging through the modelling process ("insight"). The immediate post-workshop evaluation suggested that this had occurred – participants volunteered results that had appeared in the pre-workshop evaluation of other participants, and other results that did not appear at all in the pre-workshop evaluation (see Table 3). The delayed evaluation showed that participants are more likely to retain new insights from the modelling process than ideas transferred from other participants (p<.01).

Table 3: Retention of transferred responses versus new insights

	Portion of responses

Post-Workshop Evaluation		
Retained from pre-workshop evaluation	0.08	
Transferred responses from other participants	0.41	
New insights	0.51	
Delayed Evaluation		
Transferred responses retained from post-workshop	0.16	
New insights retained from post-workshop	0.29	
Other	0.55	

"Rated preference" method

Participants were next asked to rank the answers they had volunteered in the pre-workshop and post-workshop evaluations in order of importance. Participants were blind to which actions came from which evaluation, and were not told that they had themselves been the source of the statements to be ranked.

The relative ranks were converted into values between 1 (most important action) and -1 (least important action). An "x" ranked action among "n" number of actions was scored:

$$\text{Relative ranking score} = 2\,\frac{n - x}{n - 1} - 1$$

This system was chosen so that relative weightings from participants with a different number of statements to rank could be treated comparably (Siegel and Castellan, 1988).

The post-workshop statements were ranked significantly higher than the pre-workshop statement ($p<.01$, see Table 4), and were rated higher by most participants (19 of 27). The post-workshop statements were more likely to be top-rated and less likely to be the bottom rated statements. This strongly suggests that even one year later, the workshop participants still preferred the actions identified as most important immediately after the workshop compared to what they had identified before the workshop.

The analysis for individual groups is less clear. While two workshops showed strong preferences ($p<.01$ and $p=0.02$) for the post-workshop statements, there was no significant result for the other two workshops (Group 2 slightly preferred the pre-workshop responses). There was no significant ($p<.01$) variation by age, gender education, tenure or rank.

Table 4: Participants average rating of statements, for each workshop

	Group 1	Group 2	Group 3	Group 4	Total
Average rating of pre-workshop statements	-0.28	+0.08	-0.17	-0.27	-0.19
Average rating of post-workshop statements	+0.34	-0.09	+0.17	+0.36	+0.20
Significance (t-test)	p=0.02	p=0.48	p=0.55	p<.01	p<.01

The "action list" and "rated preference" methods provide two different measures comparing participants' conclusions in a delayed evaluation with pre-workshop and post-workshop evaluations (Table 5). These two evaluation methods reveal some similarities and some differences. Using each method, there was a clear and significant preference for actions identified in the post-workshop evaluation; however the results for individual groups did not follow a clear pattern. According to the session notes, Group 3 appeared to demonstrate the greatest insight during the process, and their pre-workshop and post-workshop evaluations showed the most difference – this was not demonstrated in the delayed evaluation, where Group 3 was consistent wth other groups.

Table 5: Comparison of evaluation methods for comparing delayed views with pre-workshop and post-workshop views

| | Action list method | | Rated preference method | |
	Relative preference for post-workshop views	Significance (t-test)	Relative preference for post-workshop views	Significance (t-test)
Group 1	0.24	p=.08	0.62	p=0.02
Group 2	0.36	p=.02	-0.17	p=0.48
Group 3	0.32	p=.08	0.34	p=0.55
Group 4	0.46	p=.08	0.63	p<0.01
Total	0.32	p<.01	0.39	p<0.01

(Results greater than zero indicate preference for post-workshop views)

"Control ratings" method

Participants were asked the same question for an objective that was not the subject of their workshop, and using statements randomly sourced from the workshop for that objective. Participants had not discussed this objective in their workshop, but interviews revealed that some (10 of 30) had discussed the objectives in another context. While there was a slight preference for post-workshop statements from the other workshop (see Table 6), this was not significant (p=0.36).

Table 6. Participants average rating of statements from other workshops

	Average rating (own statements)	Preferred (own statements)	Average rating (statements from other workshop)	Preferred (statements from other workshop)
Pre-workshop statements	-0.28	7	-0.05	13
Post-workshop statements	+0.34	23	+0.05	17
Significance (t-test)	p<.01	n/a	p=0.36	n/a

Each participant's relative rating of pre-workshop and post-workshop statements was compared to their post-workshop questionnaire results (Table 7). This suggested that participants who rated the workshops as being effective (see Scott *et al.,* 2012) were

more likely to retain insights from that workshop. The post-workshop questionnaire may therefore have some predictive value in mental model change.

Table 7: Covariance between individual post-workshop ratings of workshop efficacy, and delayed preference for post-workshop statements (n=30)

	Covariance	Pearson's Coefficient
Ratings for workshop efficacy versus delayed rating of post-workshop statements	0.54	0.69

DISCUSSION

There are considerable difficulties in trying to empirically assess the impact of workshop interventions (Shadish *et al.,* 2001). This is particularly true of delayed assessments and long-term impacts, although this has been attempted in other fields (Decker *et al.,* 1988, Steinert *et al.,* 1993, Rust, 1998). These conclude that if workshops result in practice change, these practices can persist. They do not evaluate changes in attitudes or beliefs.

Participants had difficulty recalling details of the intervention. Some recollections were inconsistent with the researcher's record of the intervention. The theory of adaptive consciousness suggests that stated recollections may in fact be verbal explanations of the participants' own mental processes – that is, what they think *must* have happened, rather than actual recollections ("introspection illusion" – Wilson 2002). Nonetheless, some participants described the process accurately and in great detail.

Few participants believed that their views had been affected by the intervention. However, twelve months after the workshop, participants were still much more likely to agree with statements they had made immediately after the workshop than immediately before. If the workshop had no enduring effect, the opposite might be expected – that views would revert to the pre-workshop baseline (Gist, Stevens and Bavetta, 2006). The data suggests that the workshops had an enduring effect on participants' views on the workshop topic. The hindsight bias (Tversky and Kahneman, 1973, Fischoff and Beyth, 1975) suggests that individuals have unreliable insight into their own learning, and often assume that their current view is the one they have always held. Group model building workshops are commonly contract services purchased by the host organisation (Vennix *et al.,* 1993). If there are inconsistencies between participants' perceptions of the impact of the workshop on their views, and actual changes in their views, this may suggest that participants (clients) are not well positioned to determine the value of the service they receive.

One possible alternative explanation for why participants preferred post-workshop statements is not that they had altered participants' views in an enduring way, but rather that the statements themselves were "better" or more artfully formed, having been refined over the course of the workshop. However, when asked to review the statements from other workshops, participants did not rate the post-workshop statements significantly higher. While this suggests that participants indeed had enduring changes in their views, it does not suggest that their post-workshop views were preferable to participants who had not been part of the process, which may be a limitation in promoting the results of a group model building workshop to others who were not part of the process.

Twelve months after a five hour intervention, participants who don't normally work together were significantly more aligned than before the workshops. Many group model building practitioners conduct long-term engagements with clients (Zagonel and Rohrbaugh, 2007), rather than short episodic interventions as explored in this paper. This study should not be read as concluding that short interventions are preferable, but that mental model change is possible *even with* short interventions.

While mental model alignment has previously been identified as an outcome of group model building (Dwyer and Stave, 2007, Rouwette and Vennix, 2008, Vennix and Felling, 2009, Rouwette *et al.,* 2011), this is the first time this outcome has been demonstrated to be an enduring effect. Other long-term studies have focussed on implementation of the conclusions from the intervention (Zagonel and Rohrbaugh, 2007) or change in participant competencies through use of simulations (Cavaleri and Sterman, 1997, Thompson 2009).

There are several hypotheses for how group model building may cause alignment of mental models, including:
- Operator logic (modelling augments managers' repertoires of means-ends response options – Richardson *et al.,* 1994).
- Systems thinking (modelling increases the ability of participants to perceive generic structures and consider causal relationships – Maani and Maharaj, 2003)
- Modelling as persuasion (mental model change depends on the combination of participants' ability to process information, the quality of arguments, and persuasive content – Rouwette *et al.,* 2011).
- Boundary objects (models are a shared representation of dependencies that participants can modify, that build trust and agreement – Black and Andersen, 2012).

Further research is planned to differentiate which model best explains the evidence in these four groups.

This paper describes four group model building interventions, with a total of 30 respondents. It provides evidence that participants' views changed and became more alike during the workshop process, and that this change persisted over twelve months. Participants were typically unaware that their views had changed. Previous studies had identified insight and alignment as outcomes of group model building, but these changes had not previously be demonstrated to persist over time. The results should give greater confidence to practitioners that their interventions will continue to have an impact on participants long after their participation, even if participants are not consciously aware of the change.

Further study is needed to determine if these results are repeatable in other contexts, and whether they also apply to quantitative modelling.

ACKNOLWEDGEMENTS

REFERENCES

Black LJ, Andersen DF. 2012. Using Visual Representations as Boundary Objects to Resolve Conflicts in Collaborative Model-Building Approaches. *System Research and Behavioural Science*. 29: 194–208.

Cavaleri SA, Sterman JD. 1997. Towards Evaluation of Systems Thinking Interventions: A Case Study. *System Dynamics Review*. 13(2): 171-186.

Cavana RY, Delahaye BL and Sekaran U. 2001. *Applied Business Research: Qualitative and Quantitative Methods*. Wiley: Brisbane.

Decker DJ, Lassoie JP, Goff GR, Parrish K. 1988. Do Workshops Work? *Journal of Extension*. 26(4).

Doyle DK. 1997. The cognitive psychology of systems thinking. *System Dynamics Review*. 13: 263-265.

Doyle DK, Ford DN. 1998. Mental Models concepts for system dynamics research. *System Dynamics Review*. 14(3): 29.

Dwyer M, Stave K. 2008. Group model building wins: the results of a comparative analysis. *Proceedings of the 2008 International System Dynamics Conference*.

Fischoff B, Beyth R. 1975. "I knew it would happen" Remembered probabilities of once-future things. *Organizational Behaviour and Human Performance*. 13: 1-16

Franco LA, Rouwette EAJA. 2011. Decision development in facilitated modelling workshops. *European Journal of Operational Research* 212: 164–178.

Gist ME, Stevens CK, Bavetta AG. 1991. Effects of self-efficacy and post-training intervention on the acquisition and maintenance of complex interpersonal skills. *Personnel Pscyhology*. 44(4): 837-861.

Greenberger M, Crenson MA and Crissey BL. 1976. *Models in the Policy Process: Public Decision Making in the Computer Era*. Russell Sage Foundation: New York.

Holsti OR. 1969. *Content Analysis for the Social Sciences and Humanities*. Addison-Wesley: Reading.

Johnson-Laird P. 1995. *Mental Models 6th Edition*, Library of Congress Publications, Washington.

Maani KE and Cavana RY. 2007. *Systems Thinking, System Dynamics: Managing Change and Complexity, 2nd Edition*. Pearson Education, New Zealand.

Maani KE, Maharaj V. 2003. Links between systems thinking and complex decision making. *System Dynamics Review* 20(1): 21–48

McCardle-Keurentjes M, Rouwette EAJA, Vennix JAM, Jacobs E. 2009. Is Group Model Building worthwhile? Considering the effectiveness of GMB. *Proceedings of the 2009 International System Dynamics Conference*.

McCartt AT, Rohrbaugh J. 1995. Managerial openness to change and the introduction of GDSS. *Organizational Science*. 6: 569-584.

Richardson GP, Andersen DF, Maxwell TA and Stewart TR. 1994. Foundations of mental model research. *Proceedings of the 1994 International System Dynamics Conference*.

Rouwette EAJA, Vennix JAM, Van Mullekom T. 2002. Group model building effectiveness: a review of assessment studies. *System Dynamics Review*. 18(1): 5-45

Rouwette EAJA, Vennix JAM. 2006. System dynamics and organizational interventions. *Systems Research and Behavioral Science*. 23(4): 451-466.

Rouwette EAJA, Korzilius H, Vennix JAM and Jacobs E. 2011. Modeling as persuasion: The impact of group model building on attitudes and behaviour. *System Dynamics Review*. 27(1): 1-21.

Rouwette EAJA, Vennix JAM, Felling A. 2009. On Evaluating the Performance of Problem Structuring Methods: An Attempt at Formulating a Conceptual Model. *Group Decision and Negotiation*. 18, 567-587.

Rouwette EAJA. 2011. Facilitated modelling in strategy development: measuring the impact on communication, consensus and commitment. *Journal of the Operational Research Society* 62: 879–887.

Rouwette EAJA. 2012. Discussant's comment – does group model building work? Evidence from and comments on the paper by Videira et al. *System Research and Behavioural Science.* Published online

Rust C. 1998. The Impact of Educational Development Workshops on Teachers' Practice. *International Journal for Academic Development.* 3(1): 72-80

Scott RJ, Cavana RY, Cameron D, Maani K. 2012. Evaluation of group model building in a strategy implementation context: a New Zealand government case study, *Proceedings of the 2012 International System Dynamics Conference.*

Shadish WR, Cook TD, Campbell DT. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (2nd Edition). Wadsworth

Steigel S, Castellan NJ. 1988. *Nonparametic statistics for the behavioural sciences.* McGraw-Hill.

Steinert Y, Nasmith L, Daigle N. 2003. Executive Skills for Medical Faculty: a workshop description and evaluation, *Medical Teaching.* 25(6):666-8

Stenberg L. 1980. A modelling procedure for public policy. In Randers, J. (ed.), *Elements of the System Dynamics Method.* Cambridge University Press: Cambridge.

Tversky A, Kahneman D. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive psychology.* 5: 207-232

Vennix JAM. 1996. *Group Model Building: Facilitating Team Learning Using System Dynamics.* Wiley.

Vennix JAM. 1995. Building consensus in strategic decision-making – System dynamics as a group support system, *Group Decision and Negotiation* 4(4): 335-355.

Wilson TD. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious.* Belknap Press

Zagonel AA, Rohrbaugh J. 2007. Using Group Model Building to Inform Public Policy Making and Implementation. IN: Qudrat-Ullah, H., Spector, J.M., Davidson, P. (eds.), *Complex Decision Making: Theory and Practice.* Springer, New York