# Client Perceptions of Reported Outcomes of Group Model Building in the New Zealand Public Sector

**Rodney J. Scott · Robert Y. Cavana ·
Donald Cameron**

**Abstract** System dynamics modellers sometimes involve decision-makers in the modelling process, a method known as "group model building". Group model building has been associated with a number of different outcomes, and it is not clear which of these outcomes are important to clients. The public sector is a significant audience for group model building interventions; this paper reports on what outcomes are most valued by potential clients in the New Zealand public sector. Senior management within four government agencies identified the employees who were most likely to commission and conduct group decision processes. These individuals participated in detailed semi-structured interviews, and completed a written questionnaire, exploring the contexts in which group model building may be useful and the outcomes sought in each situation. The results suggest that, even within the public sector, the importance of a particular outcome will depend upon context. However, public servants generally appear to value trust and agreement over policy quality when conducting group-decision processes. Knowledge of the outcomes sought by potential clients helps guide the outcomes measured by researchers, and helps practitioners to tailor communication messages to clients.

**Keywords** Group model building · Group decision making · System dynamics · Clients · Outcomes · Public sector

R. J. Scott (✉) · D. Cameron
School of Agriculture and Food Sciences, University of Queensland,
St Lucia QLD, Brisbane, Australia
e-mail: rodney.scott@gmail.com

R. Y. Cavana
Victoria Business School, Victoria University of Wellington, Wellington, New Zealand

## 1 Introduction

Over almost 40 years, system dynamics practitioners have experimented with involving the client in the modelling process (Greenberger et al. 1976). These methods are now known as "group model building" (Vennix 1995, 1996). Group model building includes a range of approaches that can be broadly categorised on two axes: the level of participation (Kolfschoten and Rouwette 2006), and the use of quantitative versus purely qualitative models (Coyle 2000). In some group model building interventions, models are built by experts with some input from participants, using quantitative modelling from the outset (Kolfschoten and Rouwette 2006). In others, the model is built in workshops with or by participants, using qualitative data. In this latter group, simulation occurs only at the end of the project (Kolfschoten and Rouwette 2006) if at all (Cavana et al. 2007).

Group model building practitioners and researchers (employing a range of participative approaches) noticed that group model building resulted in changes in the behaviour of participating individuals and groups. There have been over a hundred published studies reporting on the effectiveness of group model building (Rouwette et al. 2002). These studies note a range of outcomes which in the group model building literature are considered to be "changes in the beliefs, evaluations, intentions and behaviours of participants" (Rouwette et al. 2009, p 582).

Group model building interventions are typically conducted by expert practitioners on behalf of clients (Vennix et al. 1993). While some studies refer to the client as the organisation or organisations that hired the group model building practitioner (Vennix 1995; Rouwette 2003; Thompson 2009), others refer to the individuals who make the decision to commission or purchase the practitioners services (Andersen et al. 1997; Eden and Ackermann 2004; Rouwette et al. 2009; Rouwette 2011; Rouwette and Vennix 2011; Martinez-Moyano and Richardson 2013). In the context of this study, clients are assumed to be the individuals who make purchasing decisions on the group process used. This has some similarities with the gatekeeper role described in other papers (Richardson and Andersen 1995; Luna-Reyes et al. 2006; Rouwette et al. 2011). This study also distinguishes between clients (who make purchasing decisions) and participants (who take part in the group process).

Several recent papers have explored the use of group model building in a New Zealand public service context (e.g. Cavana et al. 2007, 2014; Scott et al. 2013, 2014a, b). These report twelve outcomes associated with group model building: insight, mental model change, enduring mental model change, mental model alignment, enduring mental model alignment, communication quality, consensus, commitment to conclusion, strategy implementation, power levelling, rating of workshop conclusions by non-participants, and perceptions of workshop conclusions by non-participants. It is not clear if these outcomes are typically important to clients, or of no consequence at all.

Group model building literature suggests that the specific goals that may be emphasised or ignored will be context-specific (Zagonel et al. 2004; Rouwette et al. 2009), and implores researchers to be very clear about the goals of an intervention (Andersen et al. 1997). However, in many studies it is not clear how the measured outcomes relate to the intended outcomes (Vennix et al. 1993; Huz et al. 1997; Vennix et al. 2000; Dwyer and Stave 2008; Eskinasi et al. 2009; Rouwette et al. 2011).

Related fields, such as "soft OR", have featured reports on what their clients typically value, and suggest that a critical question for researchers and practitioners alike is understanding what outcomes clients value (Eden and Ackermann 2004). These authors described their experiences of interacting with clients, and comment on what they believe clients value, but did not present any empirical research. This study seeks to address that deficiency and thereby contribute to the evidence base for understanding what clients of group model building typically value.

An alternate view is that understanding what clients want is part of the client engagement process—that each intervention should begin with a detailed and explicit discussion with the client on the purpose of the intervention (Martinez-Moyano and Richardson 2013). Although such discussion is a component of good practice, there are advantages for researchers and practitioners of knowing *a priori* the outcomes that clients in a particular situation are likely to value. Group model building researchers need such information in determining which outcomes warrant further attention, while practitioners can tailor their initial communication with prospective clients through understanding the outcomes that are likely to be of interest.

This paper reports on research designed to derive empirical evidence on client attitudes to group decision process outcomes. There has been an increasing trend within the public service in many countries for collaborative decision-making (Ansell and Gash 2008). As a group-decision support system (Andersen et al. 2007), group model building has been applied in many public policy settings (Mingers and White 2010). This paper reports on research conducted with a sample of New Zealand public servants who were seen by their organisations as most likely to commission and conduct group decision-making processes. Their opinions were canvassed through the use of semi-structured interviews and a numerical scale questionnaire. They were asked to rate the importance of outcomes reported in group model building studies with New Zealand public servants, and also to suggest other outcomes that were important to them. The interviews discussed when and why group-decision processes would be used, and when different outcomes were important or unimportant.

The paper is structured into four sections after this introduction. The first reviews the outcomes reported in the previous papers related to this topic. The second describes the research methods. The third section reports on the results of the interviews and questionnaire. Finally, there is a discussion of what this means for group model building research and practice.

## 2 Group Model Building Outcomes

Group model building describes a range of qualitative and quantitative system dynamics methods that involve the client in the modelling process. The recent New Zealand public service case studies cited in this paper all used only qualitative tools (Cavana et al. 2007, 2014; Scott et al. 2013, 2014a, b), but similar results have been reported using quantitative methods (e.g. Vennix et al. 1993; Huz 1999; Rouwette et al. 2011; Van Nistelrooij et al. 2012).

These case studies evaluated a number of public service group model building processes, using three evaluation tools: a survey tool (Scott et al. 2014a), a pre-

test/post-test/delayed-test questionnaire (Scott et al. 2013), and semi structured interviews (Scott et al. 2014b).

The survey was based on a popular tool used in several group model building studies (Vennix et al. 1993, 2000; Rouwette 2011) that was administered immediately after participation in a group model building workshop. This was used to confirm that participants felt that the process had contributed to increased communication quality, insights, consensus and commitment to conclusions. Strategy literature reports these outcomes as being predictive of effective strategy implementation (Skivington and Daft 1991; Noble 1999; Scott et al. 2014a). Participants also compared the process to a hypothetical "normal" meeting, and believed that group model building was comparatively more effective and more time-efficient (Scott et al. 2014a).

The survey also revealed that non-managers rated the presence of an independent facilitator as important to their experience of the workshop (Scott et al. 2014a). This was related to "power levelling" (Van Nistelrooij et al. 2012), where less-powerful members are less disadvantaged in their contribution to discussion (in this study, positional rank was used as a proxy for power).

The pre-test/post-test/delayed-test questionnaire collected participants' recommendations for actions to address the problem at hand (Scott et al. 2013). This tool was administered immediately before, immediately after, and 12 months following participation in a group model building workshop. The results of this evaluation demonstrated that participants changed their mind during the workshop, and that these decision preferences persisted for at least 12 months. Because of its enduring nature, this difference was attributed to mental model change. This tool also demonstrated that participants' views became more alike (Scott et al. 2013). Mental model change that resulted in greater similarity between participants' decision-preferences was described as mental model alignment.

Participants' new decision-preferences were from two sources—some were persuaded by the views of other participants, and others developed new insights from their participation in the process. New insights from participating were more enduring that those developed through persuasion (Scott et al. 2013).

Finally, individuals who did not participate in the workshop process did not prefer the decisions made in group model building workshops to other decision alternatives (Scott et al. 2013). A meta-analysis compared the data gathered using survey tools and post-intervention interviews (Rouwette et al. 2002). This analysis revealed no difference in the outcomes reported by participants in group model building by either data collection method.

These outcomes may be interrelated. The theory of planned behaviour (Ajzen 1991) suggests that communication quality fosters insight and consensus, and insight and consensus contribute to commitment to conclusions (Rouwette 2003, Fig. 1a). Insight, consensus, communication quality and commitment to conclusions are predictive factors supporting effective strategy implementation (Noble 1999; Scott et al. 2014a). Group model building is believed to support mental model change through a combination of persuasive arguments from other participants and novel insights from the modelling process (Rouwette et al. 2011, Fig. 1b). Where group model building has been associated with long-lasting alignment of participants (Scott et al. 2013, Fig. 1c), this has been explained as related to the enduring nature of mental mod-
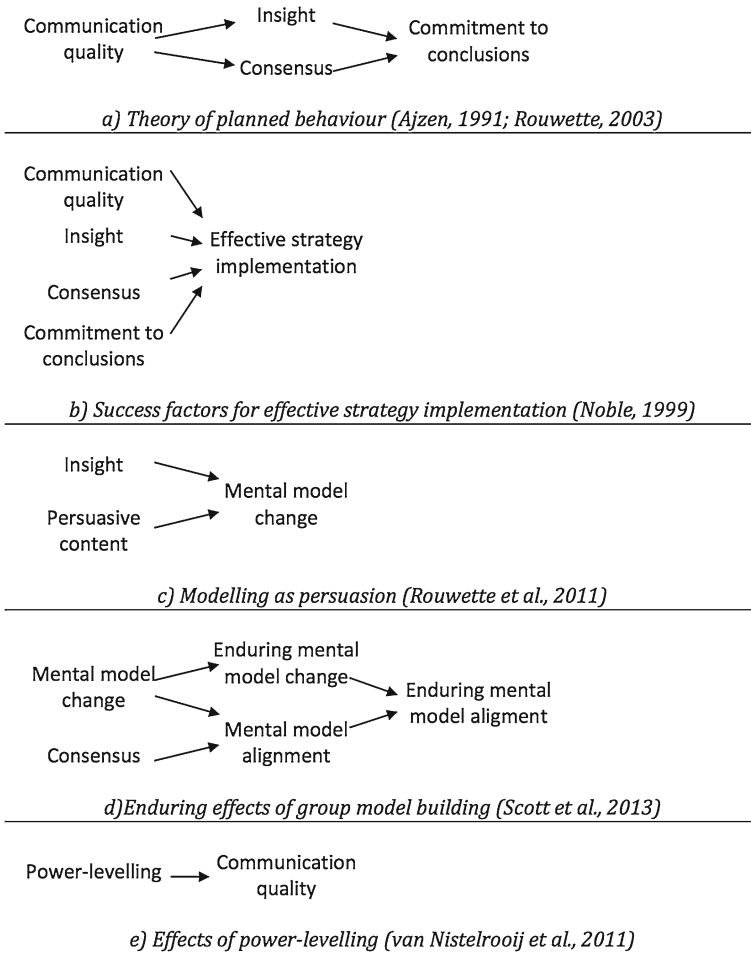
*a) Theory of planned behaviour (Ajzen, 1991; Rouwette, 2003)*

*b) Success factors for effective strategy implementation (Noble, 1999)*

*c) Modelling as persuasion (Rouwette et al., 2011)*

*d)Enduring effects of group model building (Scott et al., 2013)*

*e) Effects of power-levelling (van Nistelrooij et al., 2011)*

**Fig. 1** Theoretical relationships between reported outcomes of group model building

els of dynamics systems (Doyle and Ford 1998). Power-levelling is believed to support improved communication by providing the opportunity for more varied interactions (Van Nistelrooij et al. 2012, Fig. 1d). The rating of workshop conclusions by non-participants could not be related to the other outcomes. It is unclear how or whether these theories may be combined, though this may be an opportunity for further study.

The purpose of this study is to inform our understanding of the importance of these outcomes, and to identify other outcomes that may also be important. A definition for each outcome is included in a supplementary file.

In one of the case studies (Scott et al. 2014a), the client was asked to describe their desired outcomes for the group model building process. They indicated that they wanted to: create among employees a common understanding of their new organisational strategy; create agreed implementation actions for the strategy; and increase
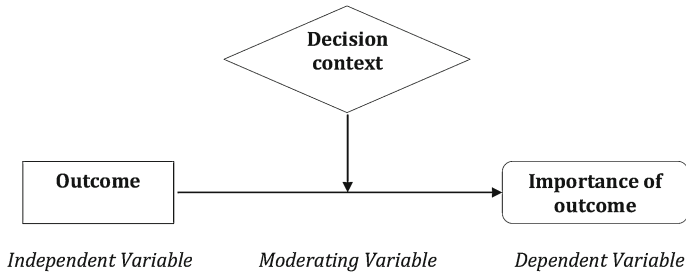
**Fig. 2** Conceptual model for the importance of group model building outcomes in the New Zealand public sector

commitment to the strategy. The prevalence of these goals is unknown, whether within other organisations, or even other problem settings (or timing) within the same organisation.

## 3 Methods

This study is a mixed methods approach to evaluation research (Blaikie 1993). Primarily qualitative methods were chosen to explore in depth the experiences and beliefs of the interviewees (Kvale and Brinkman 2008), supplemented by a quantitative survey to improve the reliability of findings (Blaikie 1993). The interviews included open questions, where interviewees identified and discussed the outcomes that were important to them, and direct questions about the reported outcomes being investigated.

The study was exploratory in nature, but the researchers hypothesised that both the nature of the outcome and several contextual factors would influence the importance of that outcome (see Fig. 2).

### 3.1 Interviews

Each research subject took part in a face-to-face interview following a semi-structured format (Kvale and Brinkman 2008). Each interview consisted of three themes: the interviewee's experiences with group-decision processes; the interviewee's desired outcomes (and when these outcomes might be most applicable); and the interviewee's opinions of the outcomes being investigated. Each of these themes is explored further below.

The interviewee was first asked to describe the context of problem-settings in which they have used group-decision processes. This included prompts on the participating parties in the group-decision process, the decision being made, and the consequence of that decision. Follow-up questions further explored the tools or processes that were used. This theme was used for three purposes: to establish the relevance of the interviewee as a person who regularly commissions or conducts group-decision processes; to investigate the kinds of problem settings encountered by public servants who use these processes; and to discover what tools were being employed.

The interviewee was then asked which outcomes were important in the experiences they had described, why these outcomes were important, and what aspects of the decision context contributed to their importance. This was used to validate later questions: in this theme, the interviewee did not know which outcomes interested the researcher, and so the opportunity for subject bias (Orne 1962, where individuals report what they think researchers want to hear) was reduced. This was also used to identify outcomes other than those being investigated.

Finally, the interviewees were supplied with each of the twelve outcomes identified in the literature (see Sect. 2). For each of these outcomes, the interviewer asked whether it was important, when it might be important, and how successful the interviewer's existing processes were in achieving this outcome. When interviewees described an outcome as sometimes important, further prompts were used to explore what factors determined whether that outcome was important or unimportant. This theme was used to evaluate each of the reported outcomes in turn.

The interviews ranged in length between 30 min and 1 h, and were recorded by an audio recorder. The interview transcripts were analysed as described in Sect. 3.4 below.

## 3.2 Questionnaire

A written questionnaire was given to the research subjects at the conclusion of the interview. The questionnaire consisted of two parts: demographic questions, and questions on the importance of each of the reported outcomes of group model building. Both are included in full as a supplementary file.

The demographic questions concerned parameters described in Table 1. Previous research had revealed that age, gender and education level had no effect on participants' reported experience of group model building (Scott et al. 2014a), but the effects of different clients' demographic variables on how they valued outcomes were unknown. If less powerful participants had previously rated the importance of an independent facilitator more highly to their experience of the process (Scott et al. 2014a), a question on organisational rank was included to determine if there was a relationship between client-rank and outcome preference.

The second part consisted of 7-point numerical scale questions to provide a quantitative indication of the importance of each of the outcomes from the literature (Cavana et al. 2001). Research subjects were asked to rate each outcome, by circling a number between 1 and 7, where 1 meant that the outcome was of no importance, and 7 meant that the outcome was very important. This provides a separate measure of the subjects' views on the different outcomes, similar to the qualitative answers in the third interview theme.

The written questionnaire was used to improve the reliability of the findings of the study. The research design was primarily qualitative, because the researchers wanted to understand the research subjects' experiences and beliefs. However, the interview questions have not been validated, so combining interview and questionnaire results in a mixed method study was used to improve reliability (Blaikie 1993). One outcome was omitted from the questionnaire in error, and this is a limitation of the study.

**Table 1** Interviewee
demographics

| Parameter | Value |
| --- | --- |
| Number of interviewees | 12 |
| Government agencies represented | 4 |
| Age (years) | |
| Mean | 44 |
| Range | 31–56 |
| Length of employment in public sector (years) | |
| Mean | 6 |
| Range | 1–20 |
| Gender | |
| Male | 6 |
| Female | 6 |
| Organisational level | |
| Director | 2 |
| Group manager | 3 |
| Team manager | 1 |
| Non-manager | 6 |
| Highest qualification | |
| Postgraduate | 9 |
| Undergraduate | 3 |
| Completed secondary | 0 |

### 3.3 Interviewee Selection

The primary researcher approached a number of New Zealand government agencies that have responsibility for developing public policy. Of these, four responded: the Ministry for Business, Innovation and Employment; the Ministry for Primary Industries; the Ministry for the Environment; and the Department of Conservation.

As discussed below, the research involved a small number of research subjects. Consequently, it was important that the subjects chosen were those who were most likely to represent the views of potential public sector clients. Hence non-probability judgement sampling methods were chosen (Cavana et al. 2001). A gatekeeper (senior executive) at each agency selected individuals in their organisation who they believed most-regularly commissioned or conducted group-decision processes, to aid work related to public policy. The researchers believed that the agencies themselves were best placed to identify the most relevant subjects for the study.

Research using qualitative interviews ideally concludes when "data saturation" has been reached; the point in data collection when no new additional data are found that develop aspects of a conceptual category (Guest et al. 2006). Conversely, experimental design frequently requires some estimate of the necessary sample size before the research has been conducted (Green and Thorogood 2009). Francis et al. (2010) propose two steps for deciding data saturation: first, specify a minimum sample size

(initial analysis sample); and second, specify how many additional interviews will be conducted without new ideas emerging (stopping criteria). The aims of the study, and characteristics of the group, influence the likely saturation point (Charmaz 2006; Mason 2010). Seven criteria have been proposed for determining an appropriate initial analysis sample size:

- the heterogeneity of the population
- the number of selection criteria
- the nesting of criteria
- groups of special interest that require intensive study
- multiple samples within one study
- types of data collection methods use
- the budget and resources available (Ritchie et al. 2003)

This study involves a selected, relatively homogenous group (public policy makers, managers, people who commission group-decision processes). There are no comparison groups, and the methods are primarily qualitative. These factors suggest a relatively small group is likely to be sufficient. Two comparable studies reported data saturation at 14 and 12 respectively (Francis et al. 2010; Guest et al. 2006).

There is no established theory on how to determine the number that should be used as stopping criteria, but three is commonly used (Francis et al. 2010). On balance, an initial sample analysis of 12 and stopping criteria of three was selected as most appropriate. After 12 interviews, the final three revealed no significant, new, unique information (i.e. data saturation was achieved). Though a robust sample for detailed qualitative study, this is a small number on which to make meaningful conclusions on the quantitative survey data – this limitation is explored further in the Discussion section. Interviewee demographics are shown in Table 1.

## 3.4 Analysis

The responses to the interview questions were transcribed, then subject to content analysis using manual coding (Cavana et al. 2001). The twelve assessed outcomes (see Sect. 2) were pre-determined as codes, as these were the main subjects of the study. Any additional outcomes mentioned by interviewees were also coded. Other codes were emergent (Holsti 1969; Strauss and Corbin 1990). The analysis was then constructed on the basis of the themes that emerged in the text, illustrated with verbatim responses where these were useful in explaining each theme.

The rated outcomes were compared using commonly applied statistical methods. The 7-point numerical scales used in the questionnaire were assumed to represent interval data (Cavana et al. 2001). A Kolmogorov-Smirnov test was used to confirm normal distribution, which allows the use of a Student's t-test to determine significance (Stephens 1974). Results for each question were compared to a neutral response (a score of 4 on the 1–7 scale), and to the overall mean (a score of 5.3 on the 1–7 scale), using a two-tailed t-test (as results could vary in either direction—Stephens 1974).

## 4 Results

Each interviewee demonstrated broad experience in commissioning and/or conducting group-decision processes, and described multiple situations where group-decision processes had been used. This confirmed that the research subjects were well selected as potential clients or users of group model building methods.

The results come from interview and questionnaire responses, and describe the importance of different outcomes in different contexts. The results were consistent with the conceptual model described earlier (Fig. 2), in that the importance of the outcome was affected by the nature of that outcome and several contextual factors. For some outcomes, interviewees described the outcome as important as a precondition to another more-desirable outcome (for example, communication quality was seen as a pre-requisite for mental model alignment). Several interviewees described outcomes as mutually reinforcing.

The importance of some outcomes was influenced by a range of contextual factors: the stage of the decision process, the participating parties in the decision, and the demographics of the client. Some outcomes were more important at different stages of the interview process, for example insight was seen as more useful in generating new ideas at the start of a process, and consensus seen as more useful at the end of a process (see Sect. 4.1). The nature of the participating parties also affected the importance of some outcomes; for example process efficiency was very important in potentially time-consuming government-stakeholder group decisions (see Sect. 4.2). Finally, client demographics had some impact on the results; while gender, age and education did not appear important, responses varied by level of experience and organisational level. (see Sect. 4.3)

The results are presented in three parts: interviewees' descriptions of the importance of each outcome; how the nature of the participating parties affected the importance of each outcome; and a statistical analysis of the questionnaire results.

### 4.1 Results for Each Outcome

Three different sources were used to determine which outcomes were most important: the second theme of the interviews, where interviewees were asked to describe the outcomes that had been important in past situations (see Table 2); the third theme of the interviews, where interviewees were asked about the importance of specified outcomes; and the written questionnaires, where respondents were asked to rate the importance of specified outcomes on a numerical scale. These three methods showed very strong agreement, with a few exceptions noted in relevant paragraphs below, where results relating to each outcome are discussed in turn. Each outcome described below is defined in the supplementary material.

*Commitment to conclusions* was the highest ranked outcome by the questionnaire responses. Interviewees distinguished between finding something acceptable for agreement in the meeting (consensus) and being committed to supporting and implementing those conclusions. Commitment was more important when the goal was to affect change (interagency cooperation, joint action with stakeholders), than when an

**Table 2**  Outcomes volunteered by interviewees as important in past group decisions

| Organisation | 1 | | | 2 | | | 3 | | | 4 | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Interview subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| Commitment to conclusions | ✓ | ✓ | – | ✓ | – | ✓ | ✓ | ✓ | – | ✓ | ✓ | – | 8 |
| Communication quality | ✓ | – | ✓ | ✓ | ✓ | ✓ | – | – | – | ✓ | ✓ | ✓ | 8 |
| Consensus | ✓ | ✓ | ✓ | – | ✓ | ✓ | ✓ | – | ✓ | ✓ | ✓ | – | 9 |
| Mental model change | ✓ | – | – | – | – | ✓ | – | – | – | – | – | – | 2 |
| Enduring change | ✓ | – | – | ✓ | – | – | – | – | – | – | – | – | 2 |
| Mental model alignment | – | – | ✓ | ✓ | – | – | – | ✓ | – | – ✓ | ✓ | – | 5 |
| Enduring alignment | – | – | ✓ | ✓ | – | – | – | – | – | ✓ | – | – | 3 |
| Effective implementation | – | – | – | – | – | – | – | ✓ | – | – | – | – | 1 |
| Persuasive content | – | – | – | – | – | – | – | – | – | – | – | – | 0 |
| Power-levelling | – | – | – | – | – | – | – | – | – | – | – | – | 0 |
| Insight | ✓ | – | – | – | – | ✓ | – | – | ✓ | – | – | – | 3 |
| View of non-participants | ✓ | – | – | – | – | – | – | – | – | – | – | – | 1 |
| Efficiency | ✓ | ✓ | – | – | ✓ | ✓ | ✓ | – | – | – | ✓ | ✓ | 7 |
| Further working together | – | – | – | – | – | – | – | ✓ | – | ✓ | – | – | 2 |
| Willingness to endorse | ✓ | – | – | – | – | – | – | ✓ | – | – | – | – | 2 |
| Attachment to language | – | ✓ | – | – | – | – | – | – | – | – | – | – | 1 |
| Participant disclosure | – | ✓ | – | – | – | – | – | – | – | – | – | – | 1 |
| Tie-breaking | – | – | – | – | – | ✓ | – | – | – | – | – | – | 1 |
| Completeness | – | – | – | – | ✓ | – | – | – | – | – | – | – | 1 |

Outcomes listed above the dotted line are the twelve outcomes investigated

agreement marked the end of the process (providing advice to a Minister or senior manager). Three interviewees mentioned that they had previously relied on voting methods to reach an agreed conclusion, however there was concern that these methods may sometimes lead to low commitment (by those whose preferred conclusions were not selected).

*Communication quality* was also highly rated by the questionnaire and interview responses. Communication quality was seen as "*crucial*" and "*where it all starts.*" In particular, communication quality was seen as important when working with stakeholders who did not have a "*shared language*" ("*engineers and planners don't even speak the same English.*"). Communication quality was seen as a prerequisite for mental model alignment which was seen as the ultimate outcome by one interviewee.

*Consensus* was generally rated as important in the questionnaire and interview responses. In many cases, coming up with "*any agreement*" was seen as success. This was particularly the case in inter-stakeholder decision processes—public servants were keen that participants all agree, even if those same convenors did not see the detail of the agreement as ideal. Several responses laboured the distinction between an ideal solution and one that all participants found acceptable for agreement. Particularly in interagency processes, participants were seen as sophisticated negotiators who

would trade off different benefits to reach an acceptable agreement (in the absence of viable alternatives to a negotiated agreement). Agreement was often achieved around non-preferred but acceptable options.

*Mental model change* was one of the lower-ranked outcomes from the questionnaire responses, but *enduring mental model change* was one of the highest ranked. Interview responses do not fully explain this difference. Mental model change was seen as a luxury by some interviewees—the goal was to reach an agreement, not have transformative experiences for the participants. Agreements were often seen as "*incremental*"— "*we're not expecting big shifts in how people see the world*". Occasionally there is a need for a "*step change*", and in those instances a technique for supporting mental model change would be desirable, but this applied to a minority of circumstances.

*Enduring mental model change* was perhaps interpreted by some interviewees as enduring agreement with the workshop conclusions; interviewees noted common delays between group-decision processes and implementation, and were particularly concerned that participants would "*go feral*" or start "*throwing stones*" at the conclusions that they had previously agreed to—"(somebody) *effectively reneging would have been a disaster*."

*Mental model alignment* was ranked moderately highly by the questionnaire responses. However, interviewees often described concepts similar to mental model alignment as their most sought-after outcomes. This was particularly true when interviewees were asked what outcomes were important to them (without being prompted with possible outcomes). Interviewees described "*shared understanding*", being "*able to understand where each other is coming from*", and "*seeing things from their point of view*" as especially important. One interviewee recalled his previous experience as a negotiator: "*People who are on opposite sides of the table don't have opposite perspectives, they have different ways of looking at the same problem*"…"*What seems a perfectly logical conclusion from your starting point, they may come to the opposite conclusion, not because they disagree with the logic but because they're coming from a different place.*" Any tools or techniques that would allow participants to see the world in a more compatible way were seen as especially desirable. From these interview responses, it might be expected that mental model alignment would have been ranked more highly among the questionnaire responses. It is possible but unconfirmed that the language "mental model alignment" was unfamiliar to respondents, and that this led to lower rankings than expected.

*Effective strategy implementation* was an outcome that did not appear well understood by some interviewees, and it was difficult to relate some answers to the questions asked. Many group-decision processes did not involve strategy implementation and therefore were not applicable. Where this was seen as important, interviewees drew distinction between talk and action ("*If you don't actually implement it, then what's the point.*") Applied business research struggles to evaluate system changes (Shadish et al. 2001), and this is an ongoing research challenge for group model building.

Some interviewees valued the *persuasive content* of the decision process used. Previous group model building research demonstrates that some learning occurs from other participants in the workshop, and some represents new ideas from the mod-

elling process (Scott et al. 2013). Interviewees were asked which of these was more important or should be more emphasised. Responses were mixed and closely followed interviewees' attitudes toward the importance of insight in their processes. Those who valued new insights saw persuasion toward existing beliefs as a barrier to creation. In contrast, those who valued agreement by any means (regardless of the quality of that agreement) saw compelling persuasion as a useful means to speed the arrival of agreement. Previous studies considering persuasion did not propose how the amount of persuasion or new insight could be increased or decreased (Rouwette et al. 2011; Scott et al. 2013).

*Power levelling* was a concept that drew polarised responses in both the questionnaire and the interviews. Having less powerful members contribute was seen as useful in generating insight ("*If it's about ideas, then you really do want to be in the situation where all participants have equal opportunity to contribute.*"), and in increasing a sense of "*engagement and ownership*" by those participants. Power-imbalances were sometimes seen as a strong barrier to participation—"*You can certainly see situations where relatively junior people are afraid to talk*" and "*you just get the loudest voices and the ones with the quickest tongues.*" Where interviewees used techniques to encourage contribution from everyone, they typically involved forcing participants to take turns in offering perspectives—interviewees talked about "*going around the room*" to elicit input individually, or using "*snowballing*" techniques to aggregate individual contributions (Thomas and Carswell 2000). This is very different to the way group model building is thought to create power levelling, through allowing contribution and modification of the model through input from all participants (Van Nistelrooij et al. 2012; Black and Andersen 2012).

In contrast, power levelling was sometimes seen as counter-productive. Toward the end of the group-decision process, "*when it comes close to closing the deal*", it was seen as sometimes beneficial for those "*who don't have authority…to sit quietly and listen to those that do.*" Some interviewees thought it represented a more durable outcome where those who had more power were more able to influence the content of the agreement—"*power is power*". Most interviewees described power levelling as relatively unimportant, and power levelling was overall rated as one of the less important outcomes of group-decision processes.

*Insight* was seen as useful "*at the beginning, to open things up*" or when "*prototyping*". However, in some cases interviewees were more interested in coming up with "*any agreement*", than whether this agreement contained any new ideas. One positive aspect of insight was that in interagency processes, new ideas were not seen as being owned by an individual agency, and so therefore were easier for other agencies to agree with. Insight was seen as unhelpful when it complicated the parameters of the discussion and delayed progress to an agreement—"*you don't want new ideas when you've trying to nail something down*." Overall, insight was not seen as very important in group-decision processes, and was the lowest ranked outcome among the questionnaire responses.

*Views of non-participants* were seen as sometimes very important and sometimes not important. In many cases, particularly where the end goal of the processes was to reach

an agreement, it was sufficient for only those present to agree, so long as those people had authority to do so ("*As long as you've got the right people in the room*"). However, in some cases described by interviewees, buy-in by broader constituencies was vital. Stakeholders were used as focus groups, with the assumption that if they agreed with a proposal it would likely be acceptable to other stakeholders with similar interests. Previous research found that conclusions developed through group model building were compelling to those present in the workshop, but not compelling to others (Scott et al. 2013). Client acceptance of solutions developed through system dynamics modelling is a long-standing challenge (Greenberger et al. 1976). Group model building aimed to overcome this challenge by involving clients in the modelling process (Vennix 1996). Where participants have to relay findings to a broader constituency, or where participants are assumed to be representative of non-participants with similar interests, the problem of compelling communication of system dynamics conclusions is resurrected. Further research is needed to develop better ways of communicating conclusions from the application of system dynamics methods (Sterman 2000).

*Efficiency* was seen as a key parameter ("*The biggest concern we have is time.*"), though participants were not specifically asked to rate its importance. Interviewees lamented that group-decision processes take considerably longer than decisions taken by individuals ("*If you were doing it by yourself, multiply the time by twenty and that's how long it takes with a group*"). Group model building participants have previously been asked to compare the speed of progress between a group model building workshop and a hypothetical "normal meeting" (Vennix et al. 1993, 2000; Scott et al. 2014a). In these studies, participants believed that group model building led to insight, consensus and commitment to conclusions more quickly than a normal meeting. If speed and efficiency are very important to public servants in designing group-decision processes, greater care should be taken in evaluating the speed of group model building processes compared to other group-decision processes.

*Further working together* was suggested by two interviewees as a key outcome of group-decision processes. In this way, participants create their own "*culture*", "*cooperation is build incrementally*", and future decisions have a foundation of mutual trust and "*goodwill*". Previous research has evaluated further use of group model building tools by an organisation (Bentham and de Visscher 1994), but not the willingness of participants to continue to work together. The boundary object mechanism for understanding group model building outcomes (Black and Andersen 2012) proposes a reinforcing loop where "our progress fuels working together". Empirical evidence of this loop would reassure public servants that use of group model building can be part of a process to build ongoing collaborative relationships.

*Willingness to endorse* was mentioned by two interviewees. This related to the inclination to publically uphold the conclusions of the decision process, and referred to situations where government was co-developing a product or programme in partnership with key stakeholders. The interviewees wanted endorsement from the group decision participants, to prevent later reputational risk to the credibility of the programme. One popular group model building research tool (the "CICC" questionnaire—Vennix et al. 1993) includes a question on willingness to endorse: "I will uphold the con-

clusions/findings of these meetings in front of other members of my organisation."
(personal communication, Etienne Rouwette 2011). If this outcome is important to
some clients, it may be useful to report specifically on willingness to endorse in future
research.

Several other outcomes were mentioned by a single interviewee only. One described
a desire for a technique to overcome participants' attachment to individual words and
to focus more on the content and meaning of the agreement—*attachment to language*
was seen as a barrier and delay to reaching agreement. This cannot be directly related
to reported outcomes of group model building. Modelling (as a visual language) may
act to interrupt any fixation on textual editing. Conversely, the act of defining vari-
ables may provide a new opportunity for language preferences to form a barrier to
agreement.

One interviewee described the need for *participant disclosure*—"*we want people
to put their cards on the table*." This can be related to two findings in the group
model building literature. In the group model building process discussed in Scott et al.
(2013), participants literally put their cards on the table—writing the variables they
believed were important on post-it notes, and sharing those with the group. Another
study investigated the extent to which unique information (information only known to
one person) was communicated within the group, and the extent to which participants
used information received (McCardle-Keurentjes et al. 2008).

Another interviewee described the need for a shortcut to decision-making between
several choices where none is obviously better. "*If you've got three (options) and
none is patently better than the others, then pick one*." The need for a mechanism for
*tie-breaking* was seen as sometimes stalling otherwise-successful projects when near
completion. It is unclear how group model building could be useful at this stage—
applying a system dynamics perspective at this time may challenge several underlying
assumptions and re-open a process that was reaching its conclusion.

Finally, one interviewee believed that it was important to ensure that no important
factors or risks had been omitted from discussion ("*How do you check you've got
all the important stuff?*"). In context, it seemed that this focus on *completeness* was
likely related to the defensibility of the decision. System dynamics practitioners may
believe that their methods are more comprehensive or holistic; however this is difficult
to measure empirically.

There was limited focus on policy quality, except indirectly (as inferred through
the interest in insight, power levelling, and completeness).

4.2 Differences Due to the Nature of Participating Parties in the Decision Process

Interviewees were asked to describe the kinds of group decisions that they commission
and/or conduct. These were then linked to different outcomes during the interviews.
The analysis of the interview transcripts revealed that the nature of the participating
parties in the group decision process influenced the importance of different outcomes,
although some outcomes were described as important or unimportant irrespective of
the participating parties. The nature of participating parties fell mostly into five cate-
gories: political decision processes; internal decision processes; interagency decision

processes; government-stakeholder decision processes; and inter-stakeholder decision processes.

*Political decision processes* typically involved agencies supporting their Ministers in negotiation with their Cabinet colleagues, or with support parties. Though public servants supported these group-decision processes by providing information, it was rare that they had any influence over the decision-support process used, and therefore could not choose to use group model building. This study was conducted from the perspective of group model building practitioners, and therefore situations where the decision process cannot be influenced are less useful for analysis; as one interviewee noted "*We can't control what they do.*"

*Internal decision processes* typically involved consensus decisions taken by peer groups within an agency. Where there was a disparity in hierarchy, decisions tended to be taken by higher-ranked employees. These involved decisions on a course of action within a policy programme, or prioritisation and resource allocation between policy programmes. These were typically convened by a member of that peer group, were either chaired by a group member or facilitated by an independent facilitator, and required consensus agreement prior to completion—"*We were going to be locked in a room until we got this sorted.*" The exception to this pattern (mentioned by two interviewees) was when a group process was convened by a higher-ranked employee, and the group's task was to arrive at a consensus recommendation—"(*The Deputy-Secretary) expects that we can come up with something…without having to bang our heads together.*" In these situations, the group included people of different rank.

*Interagency decision processes* involved employees of different agencies attempting to reach consensus agreement on a course of action, or on a joint recommendation to Ministers. Again, these were either chaired from within the group, or involved an independent facilitator. Where Ministers had demanded a joint recommendation, processes were driven to a conclusion, and often involved participants making difficult compromises. In contrast, processes to agree on a joint course of action often included alternatives to negotiated agreement—agencies could continue to operate separately if a satisfactory negotiated agreement could not be found. Partial agreements or progress toward agreement were also considered acceptable outcomes "*Sometimes it is about moving towards consensus, rather than achieving it.*" Interagency decision processes were seen as becoming more popular, with the creation of several secretariat units just to support and facilitate these discussions.

*Government-stakeholder decision processes* involved public servants working with stakeholders to reach an agreement. Typically public servants would begin the process with a tentative proposal, which would serve as the basis for negotiation—"*You never turn up with a blank sheet.*" Despite typically holding a monopoly or monopsony position, public servants were often disadvantaged by political or reputational drivers to achieve a negotiated agreement, else the initiative would be considered a failure "*There are usually win-wins, but they also know you're not going to walk away.*" Alternately, where government was contributing funding to a negotiated agreement, it was stakeholders who had an incentive to reach agreement or walk away empty handed. One example was where government would fund the production of an educa-

**Table 3** Important and unimportant outcomes for decisions involving different participating parties

| Participating parties | Outcomes mentioned as important* | Outcomes mentioned as unimportant or detrimental* |
|---|---|---|
| Internal decisions | Consensus (3), insight (2), commitment to conclusions (2), power leveling (1) | None mentioned |
| Interagency decisions | Mental model alignment (4), further working together (2), consensus (1), insight (1) | Power leveling (2), insight (1) |
| Government-stakeholder decisions | Efficiency (3), commitment to conclusions (2), willingness to endorse (2), enduring mental model change (1), consensus (1) | None mentioned |
| Inter-stakeholder decisions | Communication quality (1), enduring mental model change (1), mental model alignment (1), efficiency (1) | Insight (1), views of non-participants (1) |

*Numbers in parentheses refer to the number of interviewees who mentioned this outcome as particularly important or unimportant for decisions involving these participating parties

tional programme, if stakeholders and government could agree to the content of that programme.

*Inter-stakeholder decision processes* involved public servants acting as convenors to facilitate agreement between other parties. The aim of these processes was to arrive at consensus agreements, such that government did not need to act as a referee between competing interests. These processes were seen as increasing in popularity as they helped government avoid making contentious decisions, and were believed by interviewees to lead to less discord between opposing parties.

Interview responses commonly related the importance of each outcome to a particular decision context (as described throughout Sect. 4.1, above). For each decision context, content analysis was used to provide a simple count of how often each outcome was mentioned as particularly important or unimportant (see Table 3). In several cases, multiple interviewees described an outcome as particularly important in a decision context, notably: consensus in internal decisions; mental model alignment in inter-agency decisions; and process efficiency in government-stakeholder decisions. This last finding is of particular interest as process efficiency was not an outcome investigated.

The importance of the different participating parties was not anticipated. It may have been useful to ask separate interview questions about each type of decision group, as this would have allowed a more thorough examination of the relationship between participating parties and outcome importance. This could form the basis for further study.

## 4.3 Statistical Analysis of Questionnaire Results

The written questionnaire was primarily used to verify the conclusions of the interviews, as explored in the discussion of each outcome above. However, a comparative analysis of the questionnaire results revealed some interesting findings.

**Table 4** Ratings of the importance of each outcome, relative to neutral and mean responses (n = 12)

| Outcome | Mean | Range | Standard deviation | Difference from neutral score (1) | Difference from overall mean (2) |
|---|---|---|---|---|---|
| Commitment to conclusions | 6.3 | 5–7 | 0.78 | +2.3** | +1.0** |
| Communication quality | 6.0 | 5–7 | 0.74 | +2.0** | +0.7** |
| Consensus | 6.0 | 4–7 | 0.95 | +2.0** | +0.7* |
| Enduring mental model change | 6.0 | 4–7 | 1.04 | +2.0** | +0.7* |
| Mental model alignment | 5.8 | 4–7 | 1.03 | +1.8** | +0.6 |
| Effective strategy implementation | 5.7 | 3–7 | 1.30 | +1.7** | +0.4 |
| Enduring alignment | 5.3 | 4–7 | 0.98 | +1.3** | +0.1 |
| Mental model change | 4.4 | 3–7 | 1.31 | +0.4 | –0.9* |
| Persuasive content | 4.3 | 2–7 | 1.71 | +0.3 | –1.0* |
| Power levelling | 4.2 | 2–6 | 1.11 | +0.2 | –1.1** |
| Insight | 4.0 | 2–6 | 1.35 | +0.0 | –1.3** |

(1) "Neutral score" is a score of 4 on a 1–7 numerical scale
(2) Overall mean = 5.3, difference rounded to one decimal place
* $p < 0.05$
** $p < 0.01$

All of the outcomes assessed were rated as equally or more important than the neutral response (a score of 4 on the 1–7 scale), and some significantly more important (Table 4). This suggests that all outcomes assessed were viewed as somewhat important, and several were viewed as very important. There was a wide range of responses—only "communication quality" and "commitment to conclusions" were always rated at 5 or higher.

Outcomes were then compared against each other. Some outcomes were viewed as more important than others. "Communication quality" and "commitment to conclusions" were both viewed as significantly more important than the other outcomes, and "insight" and "power levelling" were viewed as significantly less important. Significance was determined by comparing scores for that outcome with the overall mean score (see Sect. 3.4).

The results from the numerical scale questions were also compared to each demographic field. The greatest differences were between the responses of managers (n = 6) and non-managers (n = 6), and between interviewees who had been in the public service for more than 5 years (n = 6), and those who have been in the public services for 5 years or fewer (n = 6).

There was no significant difference (p > 0.10) in the overall mean for managers (mean = 5.4) versus non-managers (mean = 5.2). Where the groups diverged was in their rating of the importance of persuasive content; this was ranked higher by managers than non-managers (5.0–3.5, p < 0.10). The researchers had considered that non-managers might place a higher value on power-levelling, as they themselves had less institutional power, but there was no significant difference between the responses of managers and non-managers for this question (4.3–4.0, p > 0.05).

It had been considered that the outcomes valued by public servants might vary through their careers. There was no significant difference ($p > 0.10$) in the overall mean for those with more than 5 years of experience (mean $= 5.2$) and those with 5 years or fewer (mean $= 5.4$). However, experienced public servants were significantly more likely to value mental model alignment as a very important outcome (6.7–5.0, $p < 0.05$). In the interviews, more experienced public servants described "shared understanding" (possibly equivalent to mental model alignment) as critically important in group decision-making.

## 5 Discussion and Conclusions

This study has several important limitations, and caution should be taken in extrapolating results to other situations. The results are likely to be most relevant for the public sector, which could be a growing market for group model building interventions. For some outcomes that were viewed as important, there is little evidence on which to determine whether group model building is relevant, and these are potentially important research gaps. Finally, what clients want from group-decision processes has important implications for how we conceive of group model building as a service. Each of these topics is explored further below.

### 5.1 Limitations

This study investigated the stated beliefs of a small number of New Zealand public servants, to determine what outcomes they value as important in group decision-making. These were then related to recently reported outcomes of group model building.

The individuals were selected by their agencies as those who most-regularly commission or conduct group-decision processes, and so are likely to be the most relevant subjects for understanding potential group model building clients in the New Zealand public sector. Twelve individuals were interviewed. For detailed qualitative research, this number proved sufficient to achieve data saturation. For quantitative research, however, the sample size is small. The quantitative data was primarily used to support the results obtained by the interviews, and should be used with caution as stand-alone measures that are representative of any broader group.

This study relies on individuals' own stated preference for different outcomes. It is possible that these do not represent individuals' actual preferences, though it is not obvious why individuals' would (for example) choose to downplay their interest in improving decision-quality through insight. It may be preferable to explored potential clients' revealed preferences (Samuelson 1938), rather than stated preferences, but collecting this data would be more challenging.

The framing of the interview as relating to "group decisions" may have led interviewees to focus on interpersonal (group) aspects. Perhaps asking instead about (for example) "solving complex problems" would have revealed greater preference for decision-quality rather than group agreement. Different outcomes are likely to be important in different settings; however group participation is one of the defining

aspects of group model building so framing the possible problems as "group decisions" did not seem inappropriate.

This study provides insights into the outcomes that are important to New Zealand public servants in commissioning and conduction group-decision processes. The results are consistent with international trends toward interagency and inter-stakeholder group decisions (Newman et al. 2004, and as explored further below), but it has not been demonstrated that these client-beliefs apply to other countries. Preferences in the private sector may vary from those in the study due to the different incentives of the commercial environment. Nonetheless, this study supports recent group model building research as applicable to potential-clients' interests.

### 5.2 A Growing Market?

Many problems faced by public sector organisations are highly complex, with multiple actors, multiple stakeholders, and conflicting outcomes (White 2002). This makes public policy questions obvious targets for the problem-solving and problem-structuring applications of system dynamics (Rose and Haynes 1999).

Two trends appear to be increasing the use of group-decision processes in the public sector. Instances of failed policy on issues that span organisational boundaries has driven demand for greater connectivity between agencies (Treisman 2007)—in New Zealand this has manifested in calls for greater interagency coordination by the "Better Public Service" initiative (State Services Commission 2011). Decisions based on consensus between stakeholders are thought to be more enduring that those arbitrated by government decision, leading to increased use of collaborative governance (Newman et al. 2004; Ansell and Gash 2008; Emerson et al. 2012)—in New Zealand this is being trialled through the consensus-based "Land and Water Forum" (Eppel 2013). This growing field lacks agreed and accepted methods for supporting group decision-making (Kim 2008; Plottu and Plottu 2011; Eden and Ackermann 2013). The opportunity for group model building in the public sector appears large, and is likely to be growing even larger (Bayley and French 2008).

### 5.3 Implications

To determine the potential of group model building to fill this opportunity, it is important to develop a sound empirical basis for the use and selection of group model building techniques. This empirical base should relate to the outcomes that potential clients are looking for.

The results of this study suggest that, in most settings, public servants who commission group decision processes are primarily interested in efficiently reaching an agreement between participants (consensus). Participants should be willing to publically endorse these agreements, and to act on them when appropriate (*commitment to conclusions*). These are areas where there is strong evidence to support the effectiveness of group model building (Vennix et al. 1993; Huz 1999; Vennix et al. 2000; Dwyer and Stave 2008; Eskinasi et al. 2009; Rouwette 2011; Scott et al. 2014a).

It is important that these agreements last. Government can move slowly, and commitment to these agreements must persist until the agreement can be put into action. While some group model building research evaluates enduring mental model change and alignment (Huz 1999; Scott et al. 2013), further research is needed to evaluate enduring agreement and the durability of commitment. It may be difficult to evaluate these outcomes due to problems of attribution (Rohrbaugh 1987; McCartt and Rohrbaugh 1989, 1995; Shadish et al. 2001).

Public servants who commission group decision processes are also interested in several outcomes for which the evidence is more limited. They are concerned by the speed it takes to reach a decision, for which group model building literature can provide only indirect evidence (participants making comparisons to hypothetical meetings, Vennix et al. 1993, 2000; Scott et al. 2014a). They are also interested in building trust and goodwill between participants, that in turn fuels future cooperation, an area that requires evaluation in group model building literature.

The lukewarm attitudes to achieving new insights were somewhat surprising, as was the general lack of interest in policy quality. Interviewees often seemed so focussed on reaching any agreement, that policy quality seemed a lesser concern. This is likely to be important as group model building practitioners think about how to describe the potential benefits of their techniques to potential customers.

The study shows that different outcomes are valuable in different contexts. The group model building literature is currently missing practical guidance on how to vary the processes used to emphasise or enhance different outcomes. Three areas of literature provide helpful but incomplete clues in this regard: experimental studies on learning outcomes; a meta-analysis of the outcomes of qualitative versus quantitative processes; and participants' own rating of the contribution of different process elements. Each is explored further below.

Several experimental studies compare the presence or absence of group model building components and how these contribute to various outcomes. These studies have evaluated the importance of the presence of a facilitator (Shields 2001; Borštnar et al. 2011), the creation of causal loop diagrams (Fokkinga et al. 2009), and the opportunity for group feedback and discussion (Škraba et al. 2003, 2007; Borštnar et al. 2011). Unfortunately these studies were conducted in experimental settings unlikely to be representative of real world behaviours (Scott 2014).

A meta-analysis found quantitative modelling processes are associated with more commitment to conclusions, consensus and system change than qualitative only processes (Rouwette et al. 2002). However, this analysis did not compare like interventions, as the quantitative processes involved far greater time commitment by participants (Scott 2014).

Other studies ask participants to rate the contribution of different components to the success of the intervention (Vennix et al. 1993, 2000; Eskinasi et al. 2009; Scott et al. 2014a). There are limitations to the ability of individuals to describe their own learning (Nisbett and Wilson 1977; Doyle 1997), and further the study design did not allow each component to be related to individual outcomes.

Further guidance is required to allow practitioners to tailor their practice toward particular outcomes.

Despite broad variance across different decision contexts, the results of this study showed generally strong support for interpersonal outcomes relating to trust and agreement, and generally less support for outcomes relating to policy quality. A similar distinction is evident in two contrasting perspectives of group model building sessions (Andersen et al. 2007). One perspective considers the model as an allegedly realistic representation of the external policy environment ("micro world"—Zagonel 2002; "virtual world"—Sterman 2000). The second perspective considers the model as a socially constructed artefact for building trust and agreement ("boundary object"— Zagonel 2002; Black and Andersen 2012; Black 2013; Franco 2013; Scott et al. 2014b; "transitional object"—Eden and Ackermann 2006). This study suggests that, in group-decision processes in the public sector, the "boundary object" perspective may be most applicable.

In conclusion, this study demonstrates that even within the public sector there exists a broad range of different group-decision contexts with different aims. In general, the research subjects preferred consensus and commitment to conclusions to cognitive change, which suggests the boundary object perspective of group model building may be most relevant to their needs. Most outcomes reported in group model building literature are valued by potential clients, but more research is required to compare the process efficiency of group model building with other methods.

# References

Ajzen I (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211

Andersen DF, Richardson GP, Vennix JAM (1997) Group model building: adding more science to the craft. Syst Dyn Rev 13(2):187–203

Andersen DF, Vennix JAM, Richardson GP, Rouwette EAJA (2007) Group model building: problem structuring, policy simulation and decision support. J Oper Res Soc 58(5):691–694

Ansell C, Gash A (2008) Collaborative governance in theory and practice. J Public Adm Res Theory 18(4):543–571

Bayley C, French S (2008) Designing a participatory process for stakeholder involvement in a societal decision. Gr Decis Negot 17(3):195–210

Bentham JB, de Visscher AG (1994) Systems thinking and its influence on operational culture. In: Proceedings of the 1994 International System Dynamics Conference. System Dynamics Society, Albany

Black LJ (2013) When visuals are boundary objects in system dynamics work. Syst Dyn Rev 29(2):70–86

Black LJ, Andersen DF (2012) Using visual representations as boundary objects to resolve conflicts in collaborative model-building approaches. Syst Res Behav Sci 29:194–208

Blaikie N (1993) Approaches to social enquiry. Polity, Cambridge

Borštnar MK, Kljajić M, Škraba A, Kofjač D, Rajkovič V (2011) The relevance of facilitation in group decision making supported by a simulation model. Syst Dyn Rev 27(3):270–293

Cavana, RY, Smith T, Scott RJ, O'Connor S (2014) Causal mapping of the New Zealand natural resources sector system. In: Proceedings of the 2014 International System Dynamics Conference. System Dynamics Society, Albany

Cavana RY, Delahaye BL, Sekaran U (2001) Applied business research: qualitative and quantitative methods. Wiley, Brisbane

Cavana RY, Boyd DM, Taylor RJ (2007) A systems thinking study of retention and recruitment issues for the New Zealand army electronic technician trade group. Syst Res Behav Sci 24(2):201–216

Charmaz K (2006) Constructing grounded theory: a practical guide through qualitative analysis. Sage, Thousand Oaks

Coyle RG (2000) Qualitative and quantitative modelling in system dynamics: some research questions. Syst Dyn Rev 16(3):225–244

Doyle JK (1997) The cognitive psychology of systems thinking. Syst Dyn Rev 13:253–265

Doyle DK, Ford DN (1998) Mental model concepts for system dynamics research. Syst Dyn Rev 14(1):3–29

Dwyer M, Stave K (2008) Group model building wins: the results of a comparative analysis. In: Proceedings of the 2012 International System Dynamics Conference. System Dynamics Society, Albany

Eden C, Ackermann F (2006) Where next for problem structuring methods. J Oper Res Soc 57(7):766–768

Eden CE, Ackermann F (2013) 'Joined-up' policy-making: group decision and negotiation practice. Gr Decis Negot. doi:10.1007/s10726-013-9375-1

Eden CE, Ackermann F (2004) Use of "soft OR" models by clients—what do they want from them? In: Pidd M (ed) Systems modelling theory and practice. Wiley, Chichester, pp 146–163

Emerson K, Nabatchi T, Balogh S (2012) An integrative framework for collaborative governance. J Publ Adm Res Theory 22(1):1–29

Eppel E (2013) Collaborative governance: framing New Zealand practice. Institute for governance and policy studies working paper, Victoria University of Wellington, New Zealand

Eskinasi M, Rouwette E, Vennix J (2009) Simulating urban transformation in Haaglanden, the Netherlands. Syst Dyn Rev 25(3):182–206

Fokkinga B, Bleijenbergh I, Vennix J (2009) Group model building evaluation in single cases: a method to assess changes in mental models. In: Proceedings of the 2009 International Conference of the System Dynamics Society. System Dynamics Society, Albany

Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP, Grimshaw JM (2010) What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychol Health 25(10):1229–1245

Franco LA (2013) Rethinking soft OR interventions: models as boundary objects. Eur J Oper Res 231(3):720–733

Green J, Thorogood N (2009) Qualitative methods for health research, 2nd edn. Sage, Thousand Oaks

Greenberger M, Crenson MA, Crissey BL (1976) Models in the policy process: public decision making in the computer era. Russell Sage Foundation, New York

Guest G, Bunce A, Johnson L (2006) How many interviews are enough? An experiment with data saturation and variability. Field Methods 18(1):59–82

Holsti OR (1969) Content analysis for the social sciences and humanities. Addison-Wesley, Reading

Huz S, Andersen DF, Richardson GP, Boothroyd R (1997) A framework for evaluating systems thinking interventions; an experimental approach to mental health system change. Syst Dyn Rev 13(2):149–169

Huz S (1999) Alignment from group model building for systems thinking: measurement and evaluation from a public policy setting. State University New York, New York

Kim J (2008) A model and case for supporting participatory public decision making in e-democracy. Gr Decis Negot 17(3):179–193

Kolfschoten GL, Rouwette EAJA (2006) Choice criteria for facilitation techniques. In: Briggs RO, Nunamaker JF (Eds). Monograph of the HICSS-39 Symposium on Case and Field Studies of Collaboration, Hawaii International Conference of System Sciences, Hawaii, 35–44

Kvale S, Brinkman S (2008) Interviews, 2nd edn. Sage Publications, Thousand Oaks

Luna-Reyes LF, Martinez-Moyano IJ, Pardo TA, Cresswell AM, Andersen DF, Richardson GP (2006) Anatomy of a group model-building intervention: building dynamic theory from case study research. Syst Dyn Rev 22(4):291–320

Martinez-Moyano IJ, Richardson GP (2013) Best practices in system dynamics modeling. Syst Dyn Rev 29(2):102–123

Mason M (2010) Sample size and saturation in PhD studies using qualitative interviews. In: Forum Qualitative Sozialforschung/Forum: Qualitative Social Research 11(3). http://www.qualitative-research.net/index.php/fqs/article/view/1428/3027

McCardle-Keurentjes MH, Rouwette EAJA, Vennix JAM (2008) Effectiveness of group model building in discovering hidden profiles in strategic decision-making. In: Proceedings of the 2008 International System Dynamics Conference. System Dynamics Society, Albany

McCartt A, Rohrbaugh J (1989) Evaluating group decision support effectiveness: a performance study of decision conferencing. Decis Support Syst 5:243–253

McCartt AT, Rohrbaugh J (1995) Managerial openness to change and the introduction of GDSS. Organ Sci 6:569–584

Mingers J, White L (2010) A review of the recent contribution of systems thinking to operational research and management science. Eur J Oper Res 207:1147–1161

Newman J, Barnes M, Sullivan H, Knops A (2004) Public participation and collaborative governance. J Soc Policy 33(2):203–223

Nisbett R, Wilson T (1977) Telling more than we can know: verbal reports on mental processes. Psychol Rev 84(3):231–259

Noble CH (1999) The eclectic roots of strategy implementation research. J Bus Res 45:119–134

Orne M (1962) On the social psychology of the psychology experiment. Am Psychol 17:776–783

Plottu B, Plottu E (2011) Participatory evaluation: the virtues for public governance, the constraints on implementation. Gr Decis Negot 20(6):805–824

Richardson GP, Andersen DF (1995) Teamwork in group model building. Syst Dyn Rev 11(2):113–137

Ritchie J, Lewis J, Elam G (2003) Designing and selecting samples. In: Ritchie Jane, Lewis Jane (eds) Qualitative research practice. A guide for social science students and researchers. Sage, Thousand Oaks, pp 77–108

Rohrbaugh J (1987) Assessing the effectiveness of expert teams. In: Munpower JL, Phillips LD, Renn O, Uppuluri VRR (eds) Expert judgment and expert systems, vol 35. Springer, Berlin, pp 251–267

Rose J, Haynes M (1999) A soft systems approach to the evaluation of complex interventions in the public sector. J Appl Manag Stud 8(2):199–216

Rouwette EAJA, Vennix JAM (2011) Group model building. In: Complex systems in finance and econometrics. Springer, New York, pp. 484–496

Rouwette EAJA, Vennix JAM, Van Mullekom T (2002) Group model building effectiveness: a review of assessment studies. Syst Dyn Rev 18(1):5–45

Rouwette EAJA (2003) Group model building as mutual persuasion. Wolf Legal Publishers, Nijmegen

Rouwette EA, Vennix JA, Felling AJ (2009) On evaluating the performance of problem structuring methods: an attempt at formulating a conceptual model. Gr Decis Negot 18(6):567–587

Rouwette EAJA (2011) Facilitated modelling in strategy development: measuring the impact on communication, consensus and commitment. J Oper Res Soc 62:879–887

Rouwette EA, Korzilius H, Vennix JA, Jacobs E (2011) Modeling as persuasion: the impact of group model building on attitudes and behavior. Syst Dyn Rev 27(1):1–21

Samuelson P (1938) A note on the pure theory of consumers' behaviour. Economica 5(17):61–71

Scott RJ, Cavana RY, Cameron D (2014) Interpersonal success factors for strategy implementation: a case study using group model building. J Oper Res Soc. doi:10.1057/jors.2014.70

Scott RJ, Cavana RY, Cameron D (2014b) Mechanisms for understanding mental model change in group model building. Syst Res Behav Sci. doi:10.1002/sres.2303

Scott RJ (2014) Group model building and mental model change. PhD Thesis, University of Queensland

Scott RJ, Cavana RY, Cameron D (2013) Evaluating immediate and long-term impacts of qualitative group model building workshops on participants' mental models. Syst Dyn Rev 29(4):216–236

Shadish WR, Cook TD, Campbell DT (2001) Experimental and quasi-experimental designs for generalized causal inference, 2nd edn. Cengage Learning, Wadsworth

Shields M (2001) An experimental investigation comparing the effects of case study, management flight simulator and facilitation of these methods on mental model development in a group setting. In: Proceedings of the 2001 International Conference of the System Dynamics Society. System Dynamics Society, Albany

Skivington JE, Daft RL (1991) A study of organizational "framework" and "process" modalities for the implementation of business-level strategic decisions. J Manag Stud 28:45–68

Škraba A, Kljajić M, Leskovar R (2003) Group exploration of system dynamics models—is there a place for a feedback loop in the decision process? Syst Dyn Rev 19(3):243–263

Škraba A, Kljajić M, Borštnar MK (2007) The role of information feedback in the management group decision-making process applying system dynamics models. Gr Decis Negot 16(1):77–95

State Services Commission (2011) Better Public Services Advisory Group Report, New Zealand Government

Stephens MA (1974) EDF statistics for goodness of fit and some comparisons. J Am Stat Assoc 69(347): 730–737

Sterman JD (2000) Business dynamics—systems thinking and modeling for a complex world. Irwin / McGraw-Hill, Boston

Strauss A, Corbin J (1990) Basics of qualitative research: grounded theory procedures and techniques. Sage, Newbury Park

Thomas P, Carswell L (2000) Learning through collaboration in a distributed education environment. Educ Technol Soc 3(3):1–15

Thompson JP (2009) How and under what conditions client learn in system dynamics consulting engagements. PhD Thesis, Strathclyde Business School, Glasgow

Treisman D (2007) The architecture of government. Cambridge University Press, Cambridge

Van Nistelrooij LPJ, Rouwette EAJA, Vestijnen I, Vennix JAM (2012) Power-levelling as an effect of group model building. In: Proceedings of the 2012 International System Dynamics Conference. System Dynamics Society, Albany

Vennix JAM, Rouwette EAJA (2000) Group model building. What does the client think of it now?. In: Proceedings of the 2000 International System Dynamics Conference. System Dynamics Society, Albany

Vennix JAM, Scheper W, Willems R (1993) Group model building. What does the client think of it? In: Proceedings of the 1993 International System Dynamics Conference. System Dynamics Society, Albany

Vennix JAM (1995) Building consensus in strategic decision making: system dynamics as a group support system. Gr Decis Negot 4(4):335–355

Vennix JAM (1996) Group model building: facilitating team learning using system dynamics. Wiley, Chichester

White L (2002) Size matters: large group methods and the process of operational research. J Oper Res Soc 53(2):149–160

Zagonel AA (2002) Model conceptualization in group model building: a review of the literature exploring the tension between representing reality and negotiating a social order. In: Proceedings of the 2002 International System Dynamics Conference. System Dynamics Society, Albany

Zagonel AA, Rohrbaugh J, Richardson GP, Andersen DF (2004) Using simulation models to address "what if" questions about welfare reform. J Policy Anal Manag 23(4):890–901