An Evaluation Strategy for the New Building Regulatory Regime

Final Version 13 October 2005

Paul Duignan[1]

Jennifer Parker

Parker Duignan Ltd
www.parkerduignan.com
www.strategicevaluation.info
paul@parkerduignan.com

This project was undertaken in conjunction with PricewaterhouseCoopers

IMPORTANT NOTE TO THE READER: This report sets out an evaluation plan structured using the REMLogic approach. This approach is now named *Outcomes Is It Working Analysis* (OIIWA). Further information and resources on OIIWA are available at www.oiiwa.org. This is a summary report, the full technical report is also available at the OIIWA site. This is a report to the New Zealand Department of Building and Housing and therefore only represents the views of the author, not necessarily the department. It should therefore not be taken as a reflection of Department of Building and Housing views at the time it was produced or at the current time regarding their evaluation strategy for the New Zealand building regulatory regime. The Department has kindly given its permission for the report to be made available as an example of the use of the REMLogic/OIIWA methodology. Organisations are encouraged to use any aspect of the OIIWA approach for their own internal business practices but are not allowed to incorporate it into software for external use. If using any aspect of OIIWA please acknowledge its use to www.oiiwa.org. The full reference to this document is Duignan, P. (2005). *An Evaluation Strategy for the New Building Regulatory Regime*. Report to the New Zealand Department of Building and Housing (DBH), 13 October 2005. (Available from www.oiiwa.org/oiiwa/documents/130pdff.html).

---

[1] Parker Duignan Ltd, contact and more information at www.parkerduignan.com and www.strategicevaluation.info and www.oiiwa.org

I. STANDARD DISCLAIMER

In preparing this report and forming our view, we have relied upon, and assumed the accuracy and completeness of all information available to us from public sources, and furnished to us by the Department of Building and Housing. We have evaluated that information through analysis, inquiry and review but have not sought to verify the accuracy or completeness of any such information or conducted an appraisal of any assets.

We will not accept responsibility to any other party other than to the Department, to whom our report is addressed, unless specifically stated to the contrary by us in writing. We will accept no responsibility for any reliance that may be placed on our report should it be used for any purpose other than that for which it is prepared.

The statements and opinions expressed in this Report have been made in good faith and on the basis that all relevant information for the purposes of preparing this Report has been provided by the Department's management, and that all such information is true and accurate in all material aspects and not misleading by reason of omission or otherwise. Accordingly, neither Parker Duignan Ltd nor its directors, employees or agents, accept any responsibility or liability for any such information being inaccurate, incomplete, unreliable or not soundly based, or for any errors in the analysis, statements and opinions provided in this Report resulting directly or indirectly from any such circumstances, or from any assumptions upon which this Report is based proving unjustified.

We reserve the right, but are under no obligation, to revise or amend our report if any additional information (particularly as regards the assumptions we have relied upon) which exists on the date of our report, but was not drawn to our attention during its preparation, subsequently comes to light.

II. TABLE OF CONTENTS

III. EXECUTIVE SUMMARY

This evaluation strategy for the New Zealand new building regulatory regime has been developed for the Department of Building and Housing (DBH) using the REMLogic[2] outcomes methodology. Developing an evaluation strategy with REMLogic consists of asking a structured set of questions about the area of activity being evaluated. Using this method has resulted in a REMLogic Structure for the new building regulatory regime. This structure identifies and sets out in a consistent format: an outcomes hierarchy, evaluation questions, potential indicators, estimated evaluation feasibility, estimated cost, potential evaluation projects and timing. In addition, this report discusses: evaluation management structure, knowledge management and risk management. This is a technical report on the REMLogic structure and the intention is that the DBH regularly updates this REMLogic Structure . A shorter summary report, *DBH Building Regulatory Regime Evaluation Strategy Summary*, has also been produced. PricewaterhouseCoopers has prepared an additional complementary report to this one which details all of the interventions which make up the new building regulatory regime. It is intended that the REMLogic structure in this report be regularly updated by DBH, **therefore if there is a later version of this REMLogic Structure the reader should use it rather than this document.**

This document is very much a technical working document and for the purposes of most readers they would be better served by reading the summary report.

---

[2] The Research, Evaluation and Monitoring Intervention Logic Outcomes Methodology is copyright to Dr Paul Duignan who developed it. If using any aspect of this approach, please acknowledge the source as www.strategicevaluation.info. This approach can be used by any organisation for its own internal business practices but it is not allowed to be incorporated into any software.

IV. INTRODUCTION

On 24 August, the Building Act 2004 was enacted. This replaced the former Building Act 1991. The new Act was intended to strengthen the regulatory environment for the building sector by regulating building work, licensing building practitioners and setting performance standards for buildings. The Act was a response to emerging issues around the quality of buildings (in particular "weathertightness" issues) and the findings of various reviews and inquires.

REMLogic methodology has been used to develop this evaluation strategy. This methodology aims to ensure that evaluation resources are most usefully allocated for the purpose of obtaining the most helpful strategic evaluation information for decision makers. This methodology attempts to push causal attribution as high as possible towards final outcomes in order to make sure that the institutions being evaluated are not avoiding the key attributional questions which stakeholders want answered. However, in those cases where it is difficult or impossible to have a clear evaluation design at the highest level, REMLogic avoids the danger of pseudo-outcome evaluations[3] by not pushing beyond what is technically feasible and affordable. By including indicator sets as an essential building block in its approach to evaluation planning, this approach is able to integrate well with existing performance measurement and monitoring initiatives.

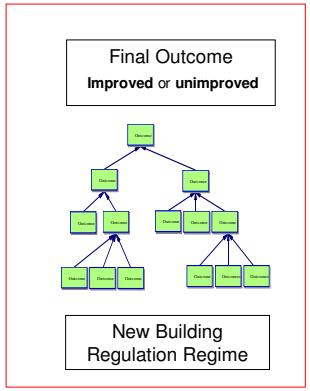
V. THE REMLOGIC APPROACH TO DEVELOPING EVALUATION STRATEGY

The REMLogic approach involves systematically working through a set of twelve questions that are discussed in detail below. Using REMLogic is an iterative process and this report sets out the initial iteration. The particular focus of this report is on identifying priority evaluation questions, and so attention has been paid to identifying indicators only in so much as they assist in identifying evaluation questions. As more work is done on identifying indicators within the DBH monitoring strategy they will be able to be incorporated within the REMLogic framework.

[3] Pseudo-outcome evaluations are evaluations which simply measure a change in outcomes and attribute this to an intervention running in the same time period without clear causal attribution.

Step 1: How is the new building regulatory regime believed to work?

To answer this, the "black box" linking the new building regulation regime to the final outcomes of improved new building work is unpacked. A logical pathway needs to be identified linking each intermediate step or outcome through which the new regulatory regime contributes right through to its final outcomes. This is referred to as an *outcomes hierarchy*[4], as generically illustrated in the figure below. Outcomes in this type of outcomes hierarchy can be seen as a "cascading set of causes in the real world".



Unpacking the "Black Box" into a cascading hierarchy of intermediate outcomes provides the basis for various types of analysis

Step 2: Is the way it is believed the new building regulatory regime works consistent with stakeholders' views, expert opinion and past evaluation findings?

[4] Another accepted name for this in the evaluation literature is an *intervention logic*. The methodology used to develop the outcomes hierarchy is set out in Duignan, P. (2004) *Intervention logic: How to build outcomes hierarchy diagrams using the OH Diagramming Approach.* www.strategicevaluation.info/se/documents/124pdff.html

To answer this, the outcomes hierarchy that has been produced should be validated against the views of stakeholders about how they think the new building regime will work, expert opinion and previous research and evaluation. It is possible that different stakeholders, including policymakers in different branches of government or in the building sector might have different perceptions of how the new building regime works and the ends that different components of the new regime are aimed at achieving. In the process of validating the new building regulation regime outcomes hierarchy against stakeholder views and expert opinion, stakeholders or experts may suggest improvements in the outcomes hierarchy.

Step 3: What routine monitoring can there be undertaken as to whether the intermediate and final outcomes are being achieved (i.e. strategic indicators)*?*

To answer this, each outcome in the outcomes hierarchy is examined to see if there are relatively easy to measure routine indicators to monitor the achievement of the new building regime's final and intermediate outcomes. At this stage the focus is simply on the strategic question of whether the outcomes that really matter are being achieved, or at least progress is being made in the right direction - not whether it can be definitely proved that it is the new regime that is changing the outcomes. Therefore such *strategic indicators* do not need to necessarily be attributable[5] to DBH interventions (however if they are attributable, that is fine). This step integrates indicator monitoring activity with evaluation planning. The figure below shows strategic indicators in the first column on the right. If strategic indicators measured over time show that final and intermediate outcomes are not being achieved, there needs to be intense critique of the activity and the wider system in which it is operating in order to see if things could be done better since the whole purpose of the activity is not being achieved (regardless of who is to blame for this).

Step 4: What routine monitoring can there be which attributes intermediate outcomes to the new building regulatory activity (i.e. attributable performance indicators)?

To answer this, each outcome in the outcomes hierarchy is examined to see if there are any relatively easy to routinely measure *attributable indicators* to monitor the achievement of intermediate outcomes. Since these performance indicators must be clearly attributable to the new building regime activity, they will tend to be at a lower level of the outcomes hierarchy than the *strategic*

[5] The reader should note again that just because a change in an outcome cannot be *demonstrably* attributed to a particular activity says nothing about whether or not the activity is actually influencing the outcome.

indicators discussed above**.** These attributable indicators should be pushed as high up the intermediate outcomes hierarchy as possible, only stopping at the point where they become impossible or too expensive to routinely measure (at which point, if they are still of any interest, they turn into non-attributable strategic indicators). In a case where feasibility and cost constraints have made it impossible to push very far up the intermediate outcomes hierarchy, the performance indicators will look like outputs – numbers of reports published, pages drafted, etc.[6]

This step integrates output level reporting with evaluation. Since these indicators can be used to hold an organisation to account for its performance, they can be referred to as *performance indicators, attributable indicators or accountability indicators.* The figure below shows the performance indicators in the second column on the right.

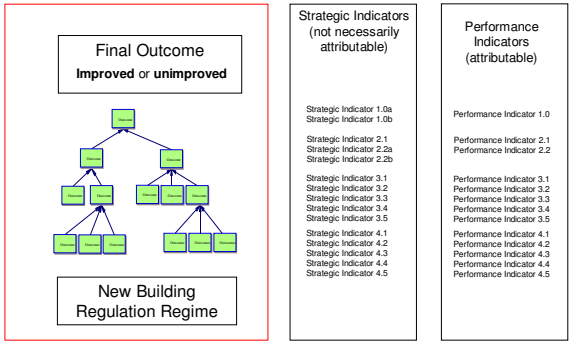
Step 5: Assess how comprehensively the strategic and performance indicators cover the outcomes hierarchy.

To do this, the coverage of indicators across the outcomes hierarchy needs to be reviewed. This process allows decisions to be made about the optimal mix of indicators that should be collected[7]. The figure below sets out a schematic presentation of the outcomes hierarchy, the strategic indicators and the performance indicators. When mapping indicators onto an outcomes hierarchy, there is likely to be both areas of indicator overlap and some blank areas within the outcomes hierarchy.

[6] Depending on the way they are defined, outputs do not have to be mutually exclusive of outcome indicator measurement and demanding mutual exclusivity creates certain technical problems. See the discussion of outcomes theory at www.strategicevaluation.info.

[7] This report was not focused on a detailed examination of indicators . Such analysis needs to be done as part of DBH monitoring strategy and can then be incorporated into the REMLogic structure.

Outcomes hierarchy, strategic and attributable performance indicators



Source: www.strategicevaluation.info

If it is found that there are no indicators in one part of the logic and a disproportionate number in another part, it may be appropriate to put more resources into developing some indicators in areas which are not currently well provided for. If there are a number of higher level indicators in either the strategic or performance indicator set then this may suggest some redundancy in indicators at a lower level within the set (because it is already known how higher level indicators are tracking). The evaluation strategy set out in this paper lists the sets of indicators that have been identified in the course of examining possible evaluation questions. It has not undertaken an assessment of the coverage of indicators across the outcomes hierarchy. It is recommended that this is done by DBH as part of its ongoing planning for monitoring the implementation of the new building regime.

Step 6: What important evaluation questions can be asked for each intermediate outcome or group of intermediate outcomes*?*

To address this, each intermediate outcome in the outcomes hierarchy is examined to see what evaluation questions could be asked about it. In addition to questions about individual outcomes, there may be some groups of intermediate outcomes or relationships between groups of outcomes for which particular evaluation questions can be asked.

Step 7: What evaluation questions are technically and practically feasible to answer*?*

The technical feasibility of answering each of the identified evaluation questions is examined. As discussed earlier, in many cases, for questions related to high-level and final outcomes, it may be impossible to answer some evaluation questions because of the nature of the activity and its context. However, **before any high level attributional evaluation questions are dismissed as not technically or practically feasible, it is essential that intense thought be put into the feasibility of answering them.** These are after all the questions that, if answered, would provide the greatest yield of strategic information. A detailed analysis of the feasibility of undertaking high-level outcome attribution evaluation is undertaken at this stage. The technical and practical feasibility of answering each evaluation question is rated into broad ranges such as: currently impossible, high feasibility, medium feasibility and low feasibility.

Step 8: What is the likely cost of answering technically and practically feasible evaluation questions?

The cost of answering feasible evaluation questions is estimated into the broad ranges: high, medium, low.

Step 9: What are priority evaluation questions within available evaluation resources?

The feasibility and cost of all of the potential evaluation questions are balanced against expected benefits of obtaining answers in order to identify those that are the highest priority for informing future decision making regarding building regulation. In prioritizing evaluation questions, the information needs of Ministers, DBH and other stakeholders should be taken into account. Step 9 ensures that evaluation spending is tightly targeted on answering priority evaluation questions for future strategy development. The power of the REMLogic methodology lies in the fact that it forces explicit, transparent and peer-reviewable decision making about exactly which evaluation questions are, and are not, being attempted. This is in contrast to evaluation strategies which may claim to be measuring "the effectiveness of the new building regulation regime" without clearly identifying exactly which evaluation questions they will be answering by mapping them onto the underlying outcomes hierarchy. In order to generate stakeholder confidence that an institution is not attempting to avoid high level attributional questions which are actually technically feasible and affordable, the priorities determined through any REMLogic exercise should be subject to rigorous peer and stakeholder review.

Step 10: How can the evaluation questions be grouped into projects and phased?

Evaluation questions which use similar methodologies, or which are closely related in terms of their implications for DBH, can be grouped into evaluation projects, and the phasing of the rollout of these evaluation projects can be decided. Evaluation projects can then be worked up into detailed evaluation plans, or Requests for Proposals (RFPs) if they are to be contracted out.

Step 11: Collecting lessons from the priority evaluation projects

Evaluation findings are mapped back onto the outcomes hierarchy as they flow in. If the outcomes hierarchy is used in strategic planning, Step 11 ensures that evaluation results are directly linked back into organisational learning and strategy rather than passing under the radar screen as scattered individual reports of individual evaluations. This step requires an underlying organisational knowledge management strategy within the DBH.

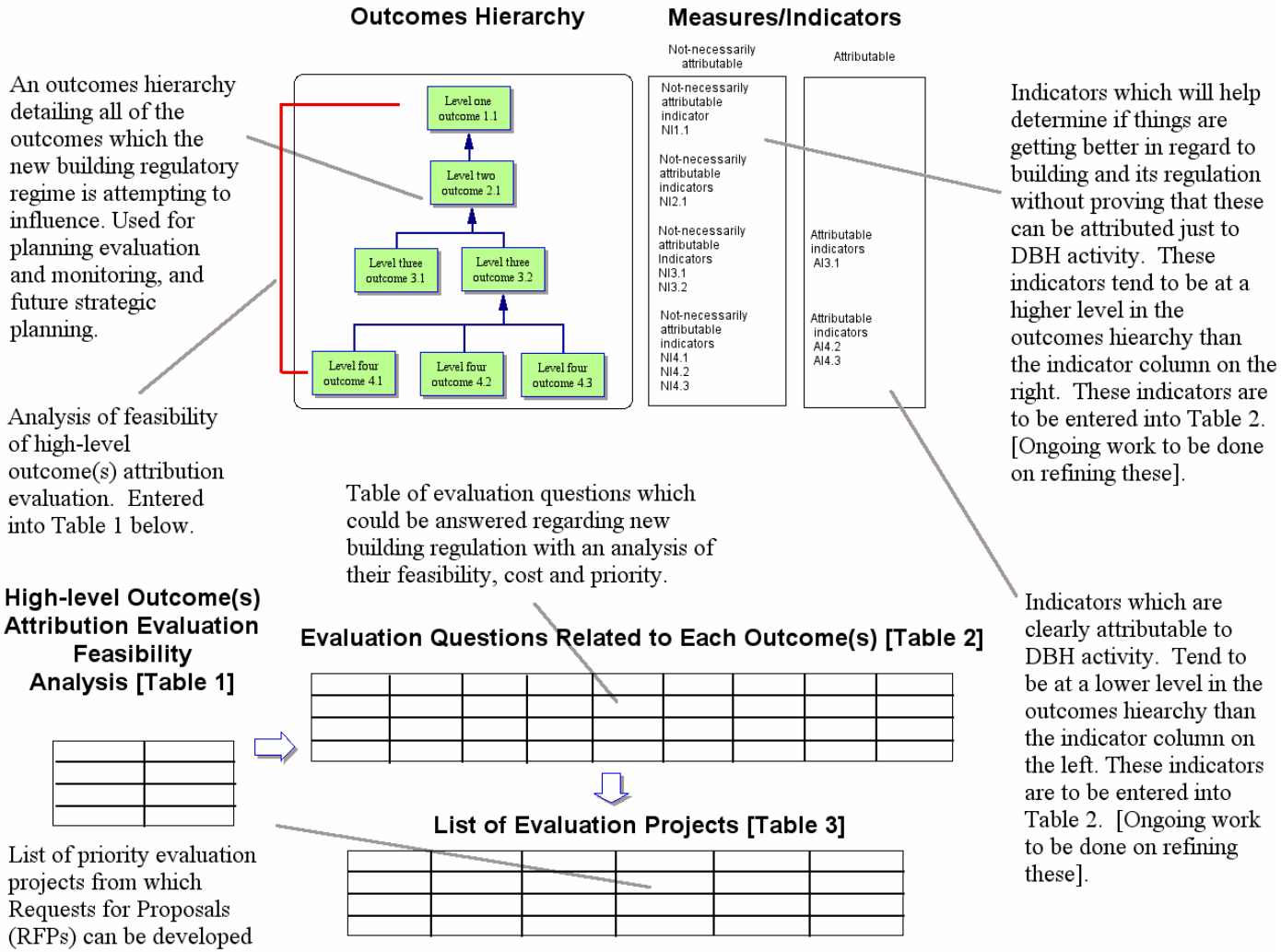
Step 12: Determining the next priority evaluation questions

The whole set of twelve questions set out here is asked again at regular intervals and a new set of evaluation priority projects is developed as the evaluation strategy evolves.

What emerges from systematically answering the questions set out above is a *REMLogic Structure* for an organization or area of activity. Such a structure is illustrated in figure below and consists of:

- a diagram of the outcomes hierarchy;
- a detailed feasibility analysis of possible evaluation designs for the high-level outcome attribution evaluation question(s) (Table 1)
- a table of evaluation questions and analysis of feasibility and cost (the table includes the outcome area, the two types of indicator, evaluation question, previous evaluation findings, feasibility, cost and priority identified) (Table 2)
- a list of priority evaluation projects (Table 3)

The New Building Regulatory Regime REMLogic Structure



VI. NEW BUILDING REGULATORY REGIME REMLOGIC STRUCTURE

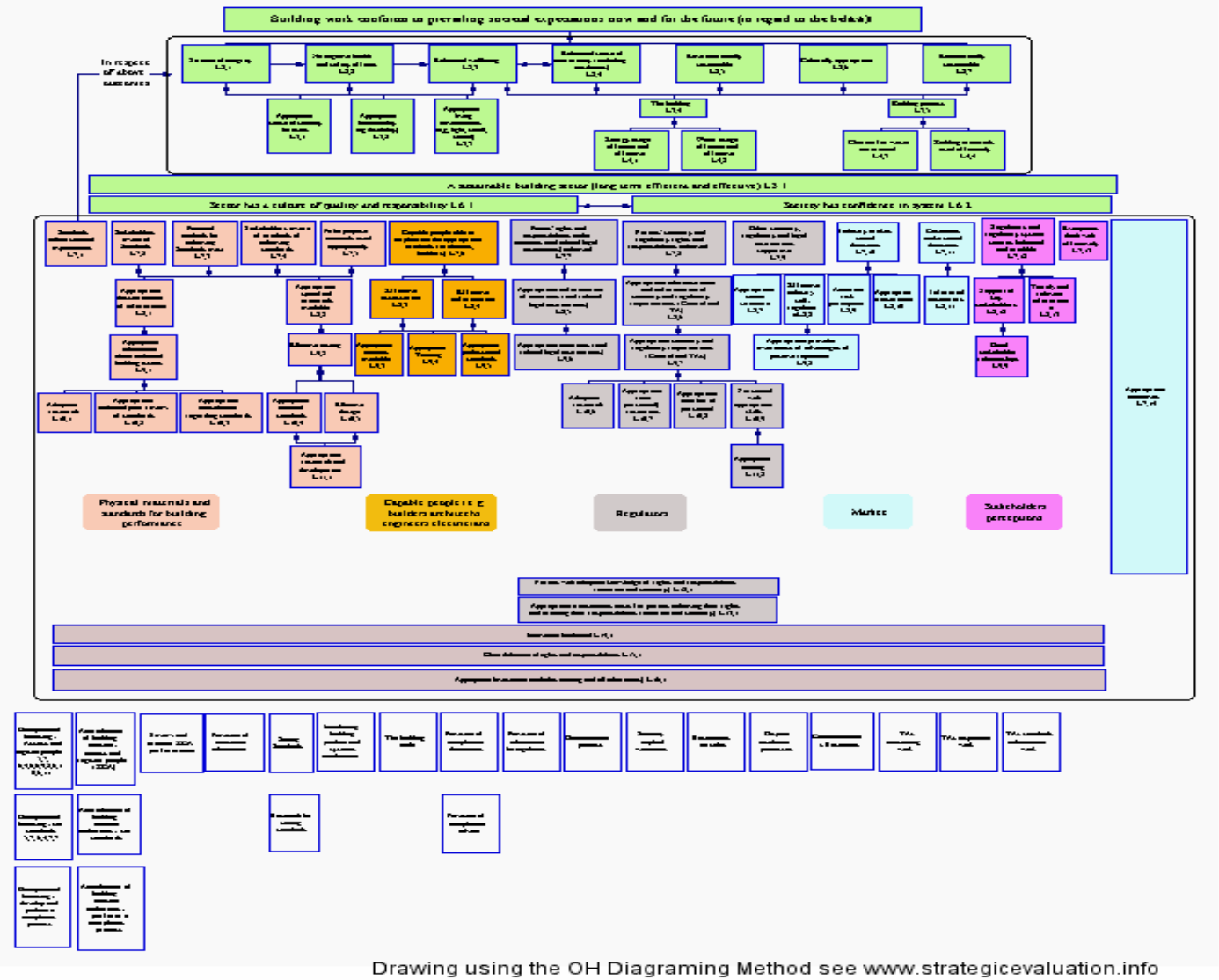
The REMLogic Structure that has been developed for the new building regulatory regime is set out below. It should be noted that REMLogic Structures are *living documents* in the sense that they should be constantly updated as the evaluation strategy evolves.

A. Outcomes Hierarchy

The outcomes hierarchy[8] for the new building regulatory regime is set out below. The first diagram shows an overview of the outcomes hierarchy and the following three diagrams show each of the three aspects of the outcomes hierarchy: the outcomes related to new building work itself; outcomes related to the regulatory regime; and a list of the interventions being used. This was developed over the course of a series of meetings in June 2005 involving DBH, PricewaterhouseCoopers and Parker Duignan Ltd staff.

[8] The outcomes hierarchy conforms to the requirement of the OH Diagraming method set out in Duignan, P. (2004). *Intervention logic: How to build outcomes hierarchy diagrams using the OH Diagramming Approach.* http://www.parkerduignan.com/se/documents/124pdff.html

The New Building Regime Outcomes Hierarchy Three Sections

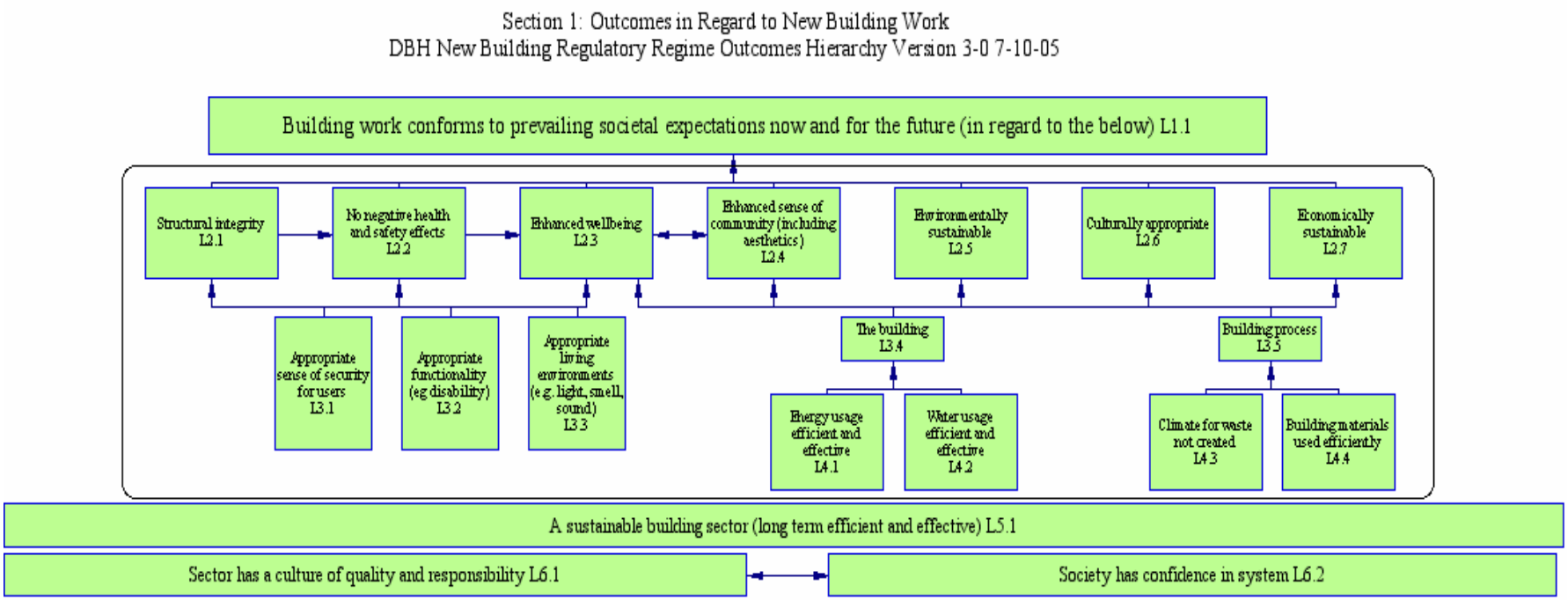


Section 1:
Set of outcomes in regard to the new building work, e.g. structurally sound, environmentally sustainable, economically sustainable etc.

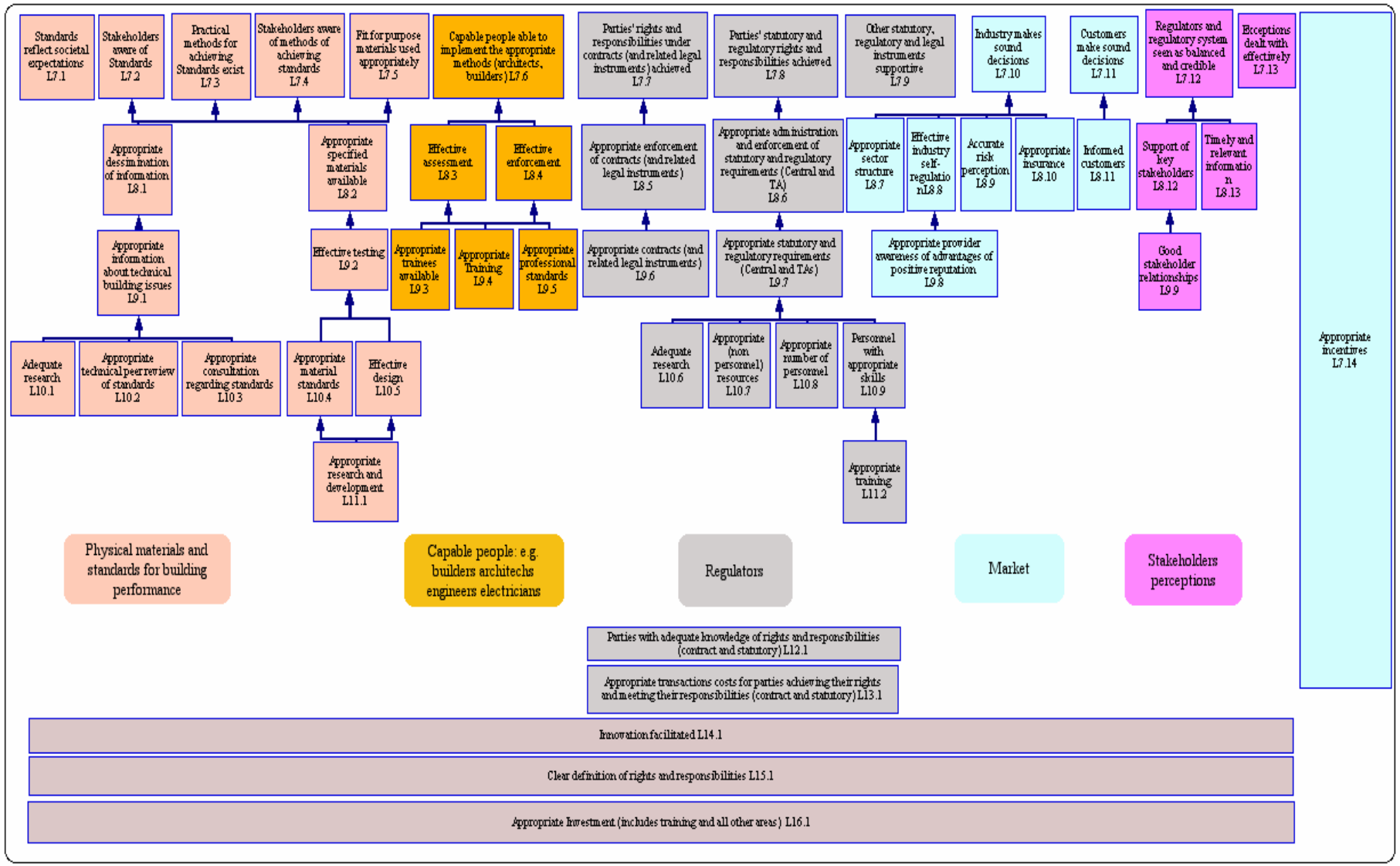
Section 2:
Set of outcomes in regard to the new building regulation system, e.g. standards, licensing, contracting, statutory rights enforced, well-functioning building markets etc.

Section 3:
Set of DBH interventions, e.g. occupational licensing, standards setting, dispute resolution process etc.

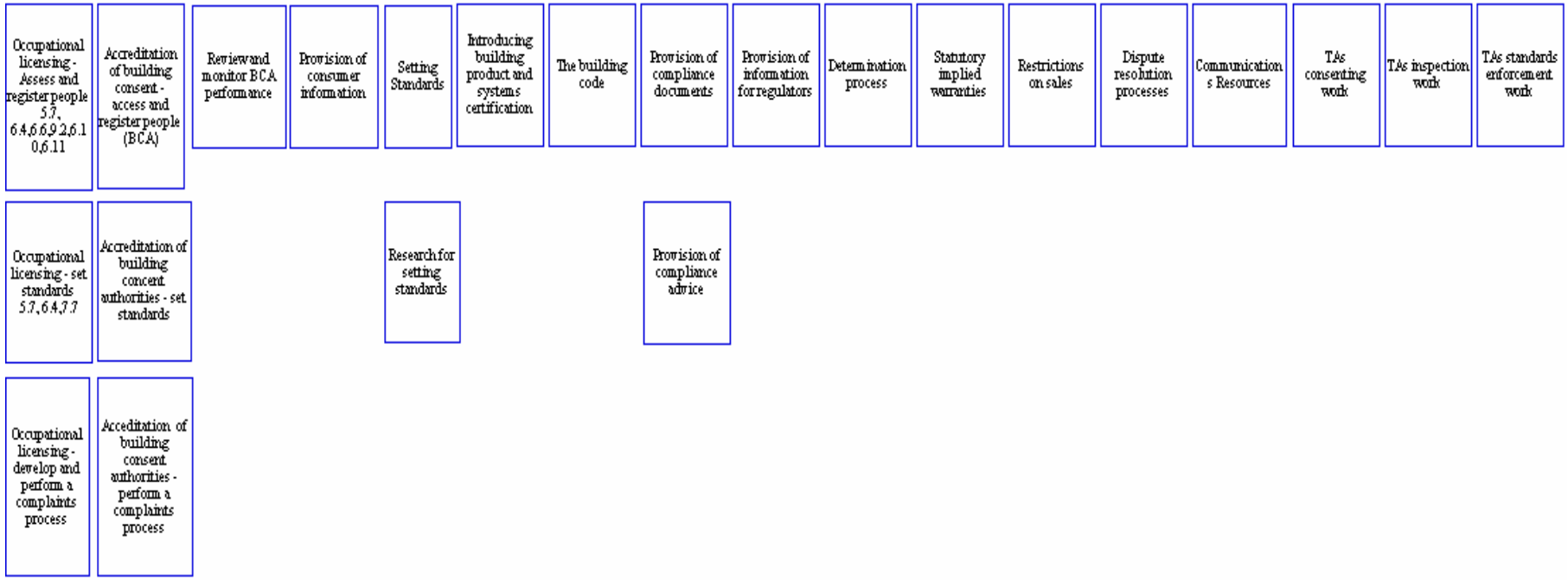
Drawing using the OH Diagraming Method see www.strategicevaluation.info


Section 1: Outcomes in Regard to New Building Work
DBH New Building Regulatory Regime Outcomes Hierarchy Version 3-0 7-10-05
Building work conforms to prevailing societal expectations now and for the future (in regard to the below) L1.1
Structural integrity L2.1
No negative health and safety effects L2.2
Enhanced wellbeing L2.3
Enhanced sense of community (including aesthetics) L2.4
Environmentally sustainable L2.5
Culturally appropriate L2.6
Economically sustainable L2.7
Appropriate sense of security for users L3.1
Appropriate functionality (eg disability) L3.2
Appropriate living environments (e.g. light, smell, sound) L3.3
The building L3.4
Building process L3.5
Energy usage efficient and effective L4.1
Water usage efficient and effective L4.2
Climate for waste not created L4.3
Building materials used efficiently L4.4
A sustainable building sector (long term efficient and effective) L5.1
Sector has a culture of quality and responsibility L6.1
Society has confidence in system L6.2

Section 2: Outcomes in Regard to New Building Regulation System
DBH New Building Regulatory Regime Outcomes Hierarchy Version 3-0 7-10-05



Section 3: Set of DBH Interventions
DBH New Building Regulatory Regime Outcomes Hierarchy Version 3-0 7-10-05



B. Evaluation Questions Table

The tables below set out the intermediate outcomes from the outcomes hierarchy, potential strategic (not necessarily attributable) indicators, potential attributable indicators, previous evaluation findings, potential evaluation questions, feasibility (technical and practical) of answering the question, cost and priority. The first table covers the main overall evaluation question which is whether the new New Zealand building regulatory regime resulted in a change in the quality of new building work in New Zealand. Due to the importance of this evaluation question, the feasibility of various methods of answering it are discussed in the Main Evaluation Question Outcomes Methodology Options Analysis Table. This is then followed by a tabular analysis of all of the other identified possible evaluation questions.

1. The Main Overall Evaluation Question

The table below sets focuses just on the main overall evaluation question and is followed by outcome evaluation options analysis.

(Inter - mediate) Outcome	Strategic indicators (may not be attributable)	Attrib- utable indicat- ors	Previous evaluation findings	Evaluation questions [EQ]	Method & feasibility (technical and practical)	Cost[9]	Priority

[9] These are only very rough initial estimates of cost for the purposes of initial planning. They should be subject to peer review and then further consideration by DBH before they are acted upon in any way.

L1.1	No separate indicator currently measured or planned. To develop an independent indicator for this would require a regular representative survey of new building work from around the country. Already all building work is intensively inspected by Territorial Authorities (TAs) as part of the regulatory regime and this will be subject to more intensive accreditation audit in the future by DBH as part of the new regime. So as long as the regulatory regime is working (which will be measured by lower level indicators and evaluative activity) it will be assumed that building standards are, by definition, improving as if new building does not conform to the new requirements it is not allowed to be built. It may be possible to use the number of complaints or negative publicity tracking in this area.	None	The Hunn report suggested the previous building regulatory regime failed to ensure adequate improvement in the quality of new building in New Zealand	EQ1: Has the new building regulatory regime resulted in new building work conforming to prevailing societal expectations?	Not feasible [see the Main Outcome Evaluation Methodology Options Table below for details]		
				EQ2: In the opinion of an independent expert(s) has the new building regulatory regime contributed new building work conforming to prevailing societal expectations?	Highly feasible [see the Main Outcome Evaluation Methodology Options Table below for details]	$50,000	High

It is essential that all possible methodologies for answering the high-level outcome evaluation have been considered. The table below sets out thinking to date on all possible outcome evaluation designs. In a REMLogic structure, outcome evaluation designs can be divided into six possible designs as set out below.

EQ1: Has the new building regulatory regime resulted in new building work conforming to prevailing societal expectations?

Experimental design - setting up a comparison between a group which receives the intervention and a group (ideally randomly selected from the same pool) which does not.	Not feasible	For ethical, political, legal, and design compromise reasons it is not possible to implement the interventions in one or more localities while other localities (serving as a control group) do not have the interventions. Apart from anything else, statutory regulation could not be imposed only part of the country. In addition, there is a major design compromise problem given the practical and political importance of having a high standard of new building work it is likely that *compensatory rivalry* would reduce any difference in outcomes between the intervention and control group;[10]
Regression discontinuity design - graphing those receiving the intervention on a measurable continuum (e.g. research excellence as measured from annual reports) and only applying the intervention to those below a certain cut off level, the effect should be apparent by an upwards shift of the graph at the cut off point.	Not feasible	In theory it would be possible to rank local authorities in order of the quality of their new building work and if resources for the intervention were limited then it would be ethical to only intervene in those with the worst new building going on and hence establish a regression discontinuity design. However, the political, legal and design compromise problems (as in the above design) mean that a regression discontinuity design is not feasible.
Interrupted time series design -	Not feasible	This design would be possible if multiple measures of new building quality were available over a

[10] Compensatory rivalry is where the control locality also implements the interventions which are being evaluated because it also wants to achieve the outcomes which are as important to it as to the locality receiving the intervention. See Shadish, W.R., Cook, T.D. & D. T. Campbell (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin. p. 80.

tracking a measure of an outcome a large number of times (say 30) and then looking to see if there is a clear change at the point in time when the intervention was introduced		lengthy (say 20 years) time series which could then continue to be tracked over the course of the intervention. However this design has the design compromise problem that there is another major factor – the crystallisation of liability – which is concurrent with the introduction of the new building regulatory regime. It should be noted that this, of course, does not mean that any available time series data cannot be used as a way of tracking the strategic indicator of quality of new building work over time. It is just that any such time series would be silent on the question of attribution of change to the new building regulatory regime.
Constructed matched comparison group design - attempting to locate a group which is matched to the intervention group on all important variables apart from not receiving the intervention	Not feasible	This would require the construction of a comparison group not subject to change in its regulatory regime, ideally over the same time period as the intervention. Since the new building regulatory regime is a national intervention such a comparison group will not be able to be located in New Zealand. It is theoretically possible that one or more comparison groups could be constructed from other countries or localities within other countries. However discussion with DBH staff concluded that it is virtually impossible for the assumptions of such a design to be met. These assumptions are: that the initial regulatory regime in the other country was the same as that in New Zealand; that the conditions new buildings are exposed to in the other country are similar to New Zealand; that the authorities in the other country do not respond to new building quality issues by changing the regulatory regime as has been done in New Zealand; and that there is a sufficiently valid and reliable way of measuring new building quality in both countries before and after the intervention. It should be noted that while some of these assumptions may be met in regard to some overseas countries, all of them would need to be met for a particular country for it to provide an appropriate comparison group.
Causal identification and elimination design - detailed analysis of all of the possible causes of a change in an outcome and elimination of all other causes apart from the intervention	Low feasibility	In some cases it is possible to develop a detailed list of possible causes of observed outcomes and then to use a forensic type process to identify what is most likely to have created the observed effect. This goes far beyond just accumulating evidence as to why it may be possible to explain the observed outcome by way of the intervention and requires that the alternative explanations be eliminated as having caused the outcome. This may not be possible in this case due to the concurrent crystallisation of liability which occurred in the same timeframe as the intervention. It is likely that this cause is significantly intertwined with the intervention in being responsible for any change that occurs in new building practice and that it will be impossible to disaggregate the effect of the intervention from the effect of crystallisation of liability. Further work should go into ensuring that this design option is not feasible.

Expert connoisseurship design - asking subject expert(s) to analyse a situation in a way that makes sense to them and to assess whether on balance they accept the hypothesis that the intervention may have caused the outcome.[11]	High feasibility	One or more well regarded and appropriate independent expert(s) in building regulation (presumably from overseas in order to ensure independence) could be asked to visit New Zealand and to assess whether they believe that any change in new building outcomes is a result of the new building regulatory regime. This would be based on their professional judgement and they would take into account what data they believe they require in order to make their judgement. Their report would spell out the basis on which they made their judgement. This approach is highly feasible but provides a significantly lower level of certainty than all of the other outcome evaluation designs described above. If this design is used then the evaluation question being answered would always have to be clearly identified as: *In the opinion of an independent expert(s) has the new building regulatory regime contributed new building work conforming to prevailing societal expectations?* There are obvious linkages between this design and the causal identification and elimination design above and the feasibility study for that design should also look in detail at the possibilities for the expert connoisseurship design.
Stakeholder judgement design – asking stakeholders to judge a situation in a way that makes sense to them and to assess whether on balacne they accept the hypothesis that the intervention may have caused the outcome.[11]	High feasibility	A selection of stakeholder key informants (key informants are people who have knowledge of what has occurred in an intervention) could be interveiwed in face to face interviews and their opinions regarding what outcomes can be attributed to the new building regime could be summarised and analysed in order to draw general conclusions about the effect of the intervention. This could be linked in with an expert connoisseurship and causal elimation design as described above.

[11] The last two designs would not usually be expected to establish causality as robustly as the other listed designs. However these designs are frequently used and deserve a place in a full typology of outcome evaluation designs; in particular circumstances they are feasible, affordable and accepted by stakeholders as better than having no high-level outcome attribution information. Even though they are often more feasible and affordable than the other designs, decision-makers have to consider on a case by case basis whether these designs can actually provide any coherent information about attribution or whether they will just end up being examples of pseudo-outcomes studies which do not contribute any sound information about attribution.

2. Other Lower-Level Evaluation Questions

The table below analyses the lower level evaluations questions which have been identified from discussions with DBH. It should be noted that these are evaluation questions in regard to the **new building regulatory regime**, not in regard to evaluating the DBH as a whole.

Inter - mediate Out-come	**Potential strategic indicators (may not be attributable)**	**Potential attributable indicators (clearly attributable to activity)**	**Previous evaluation findings**	**Potential evaluation questions**	**Method & feasibility (technical and practical)**	**Cost**	**Priority**
Evaluation plan	N/A	N/A	N/A	EQ3: Is the evaluation plan sound and can it be improved?	High feasibility	Low	Highest
Whole outcomes hierarchy	N/A	N/A	N/A	EQ4: Is the outcomes hierarchy a comprehensive and well structured set of all of the important intermediate outcomes which need to be achieved?	High feasibility	Not significant	
Outcomes evaluation methodology	N/A	N/A	N/A	EQ5: Can an outcomes evaluation methodology be designed based on a *causal identification and elimination design* and linked to an *expert connoisseurship design?*	High feasibility for at least developing a design which relies on the expert connoisseurship design.		

L7.1 – L16.1	Same as attributable indicator	A basket of indicators will be available flowing off the new building regulatory regime (e.g. accreditation results in regard to BCA/TAs). The improvement in the building regulatory regime can be seen as being reasonably attributable to DBH. Tracking these indicators over time will be able to provide an indication of improvement in the building regulatory regime.	The results from the Hunn report provide a baseline for this	EQ6: In the opinion of an independent expert(s) has the new building regulatory regime improved in quality over time?	High feasibility. A Hunn report type of review of the regulatory regime can be repeated twice more to review whether the regulatory regime is improving over time.	Moderate (say $50,000)	High
				EQ6a: Can a comprehensive but concise set of indicators be developed that will allow monitoring of the new building regulatory regime?			

L7.1 Standards reflect societal expectations	No overall indicator	No indicator overall indicator although some lower level indicators may provide some indication of this	Sector consultation processes. *Open space technology* workshop looking at expectations. Whole of the standards development process.	EQ7: Is the building code reflecting prevailing societal expectations?	Feasible but only by replicating existing DBH processes. The DBH already undertakes a number of processes directed at ensuring that standards reflect prevailing societal expectations. Therefore this question will not be answered separately from the DBH process (however see EQ8 below)	Potentially high cost to simply replicate existing DBH processes	Low
				EQ8: Does the DBH have a sound process for ensuring that the building code reflects prevailing societal expectations?	High feasibility	Part of process evaluation of DBH processes. Project total $150,000 say	Medium

L1.1-L4.4	Same as attributable indicator	Given the process of inspection of all new building work by TA's and TA's accreditation for undertaking those inspections. If lower level indicators are showing improvements in this system, then for indicator purposes it will be assumed that this outcome is being met by definition. This seems to be an indicator which is reasonably attributable to DBH.		EQ9: Is new building work being undertaken in accordance with the code?	High feasibility. It is highly feasible to undertake a representative survey of new building work throughout the country to find out if it is being built in accordance with the code. Given the intensive activity of TA's inspecting all significant new building work and the accreditation regime they are going to be operating under, re-examining this issue is not justified because of the cost.	Prohibitively high	Low
L2.7	None	None	Policy analysis relevant to cost benefit analysis undertaken at introduction of new building regime	EQ10: What is the net benefit of the new building regulatory regime?	Low feasibility. Some information from the two questions below on standards.	N/A	

	None	None	Analysis undertaken prior to setting of new standards	EQ11: What is the net benefit of new building standards?	Medium feasibility. Need to factor in both initial and *whole of life* cost and benefits, building related benefits and lowered cost of later disputes etc (more difficult).		High
	Same as attributable indicator	Compliance cost figures from BCA/TAs		EQ12: What is the compliance cost of ensuring that new building work meets the new standards?	High feasibility. Routine figures available from BCA/TAs	Low	High
	Same as attributable indicator	Some international comparative figures may be available (unlikely to be comparative)		EQ13: How does the New Zealand compliance cost compare internationally?	Low feasibility. There are many difficulties in identifying overseas jurisdictions which are comparable in terms of regulatory requirements, building methods, building sector, regulatory history and climate. However it would be worth undertaking a feasibility study of whether anything robust could be done on this topic.	Medium	High
L14.1	None	Some indicator information available, e.g. alternative solutions from TAs, new product verification.		EQ14: What is the impact of the new building regime on innovation?	High feasibility	Part of process evaluation of DBH processes. Say project total $150,000	High

Whole regulatory system (L5.1 – L16.1]	None	BCA/TA accreditation information will provide some indicator information in regard to this		EQ15: Is one national approach appropriate for all regions? [to discuss]	High feasibility	Part of process evaluation of DBH processes. Say project total $150,000	Medium
Whole regulatory system (L5.1-L16.1]	None	None		EQ16: What can be learnt from other jurisdictions for improving the system?	High feasibility	Low. Regular review by DBH staff of international developments	High
L7.7, L7.8, L7.11, L7.12, L16.1	Some data could be collected from BCA/TAs	None		EQ17: Are customers satisfied with new building work?	High feasibility	Low (if collected from BCA/TAs)	
L7.12	None	None		EQ18: Is the regulatory regime seen as balanced and credible?	High feasibility	Part of process evaluation of DBH processes. Say project total $150,000	

L11.1, L10.4, L10.5 L9.2, L8.2, L7.5	Indicator information collected from materials certifiers regarding number and type of materials certified	None [to discuss]		EQ19: Is the materials certification system working effectively?	High feasibility	Moderate (say $50,000)	High
L9.3, L9.4, L9.5, L8.3, L8.4, L7.6	Indicator information collected from building practitioner licensing system	Number of hits on register web site.		EQ20: How well is the building practicioner licensing system working?	High feasibility	Moderate (say $50,000)	High
L11.2, L10.6, L10.7, L10.8, L10.9, L9.7, L8.6, L7.8		All of the information available about TA and BCA activity.		EQ21: How well are TAs and BCAs functioning as part of the regulatory system?	High feasibility	Low (say $20,000). May not cost this much, may be able to be done from the auditing and other processes DBH is putting in place.	High
Whole regulatory system (L5.1-L16.1]		Information from all system indicators informs this	Hunn report and earlier analysis leading to new building regime	EQ22: How can DBH processes be improved?	High feasibility	Conclusions drawn out of process evaluation of DBH processes. Say project total $150,000	High

C. Evaluation Projects List

Evaluation Project (EP)	Evaluation questions	Way of proceeding	Timing	Estimated cost[12]
EP1: Peer review of this evaluation plan	EQ3: Is the evaluation plan sound and can it be improved?	Send the evaluation plan to two evaluation specialists for peer review	Commissioned: July 2005 Completed August 2005	Below $5000
EP2: Stakeholder validation of outcomes hierarchy	EQ4: Is the outcomes hierarchy a comprehensive and well structured set of all of the important intermediate outcomes which need to be achieved?	1) Send the outcomes hierarchy out to selected sector key informants and ask for written or telephone comment 2) Convene a focus group of sector key informants (say up to ten sector key informants if they can be attracted to come to such a meeting) at the same time they could have the opportunity to make any initial response to the evaluation strategy plan	Undertaken July 2005	Not significant
EP3: Evaluation outcomes options design feasibility project	EQ5: Can an outcomes evaluation methodology be designed based on a *causal identification and elimination design* and linked to an *expert connoisseurship design?*	Small project involving someone with evaluation expertise to think through the possibilities. The recent work of the evaluator Michael Scriven may be helpful as a starting point for this project.[13] This project to include developing the Terms of Reference for such a study.	Commissioned: August 2005 Completed October 2005.	Below $10,000

[12] These are only very rough initial estimates of cost for the purposes of initial planning and should not be taken as any more than that. They should be subject to peer review and then further consideration by DBH before they are acted upon in any way.

[13] More information can be obtained from Dr Paul Duignan paul@parkerduignan.com.

Evaluation Project (EP)	Evaluation questions	Way of proceeding	Timing	Estimated cost[12]
EP4: Independent expert(s) view of contribution of new building regime to outcomes	EQ2: In the opinion of an independent expert(s) has the new building regulatory regime contributed to new building work conforming to prevailing societal expectations?	The exact nature of this evaluation project will depend on the findings from EP3 looking at the cross-over between this design in this case and a causal identification and elimination design. At its simplest, it would just involve asking an independent expert or experts, probably from overseas, to answer evaluation question EQ2, taking into account what data they believe they require in order to make their judgement. Their report would spell out the basis on which they made their judgement.	Commissioning: December 2005 Initial site visit: March 2006 Final site visit: March 2009	$75,000-$150,000
EP5: Replication of Hunn review	EQ6: In the opinion of an independent expert(s) has the new building regulatory regime improved in quality over time?	This project would consist of a review like the Hunn review in 2005 and in 2009.	[To be considered]	Below $100,000 depending on whether this project could be linked to project EP4 above.
EP6: Evaluation of DBH processes	EQ8: Does the DBH have a sound process for ensuring that the building code reflects prevailing societal expectations? EQ14: What is the impact of the new building regime on innovation? EQ15: Is one national approach appropriate for all regions? [to discuss] EQ18: Is the regulatory regime seen as balanced and credible?	Process evaluation using document analysis, questionnaires and key informant interviews to provide detailed examination of DBH processes.	[To be considered]	Up to $150,000 depending on whether this project could be linked to project EP4 above.
EP7: Indicator development project [if there is already an	EQ6a: Can a comprehensive but concise set of indicators be developed	Identifying both not-necessarily attributable indicators and attributable indicators, mapping them onto the outcomes hierarchy to identify how complete coverage there is, working out protocols for routine collection and	[To be considered]	Initially undertaken within DBH staff resources. Likely to require additional

Evaluation Project (EP)	Evaluation questions	Way of proceeding	Timing	Estimated cost[12]
indicator development project within DBH this would be the same project]	that will allow monitoring of the new building regulatory regime?	analysis of these indicators		funding. Say $10,000-$20,000.
EP8: Formative evaluation project	EQ22: How can DBH processes be improved?	Formative evaluation conclusions drawn from EP6 above	[To be considered]	Included within the cost of EP6 above.
EP9: Cost benefit analysis of new building standards	EQ11: What is the net benefit of new building standards? EQ12: What is the compliance cost of ensuring that new building work meets the new standards?	Cost benefit analysis to be undertaken.	Commissioned: August 2005 Phase one: cost benefit analysis framework established December 2005 Phase two: recalculation based on compliance costs 2007 Phase three: reworking if any standards change (as required) [to discuss]	$40,000-$80,000
EP10: Feasibility study of international compliance cost estimation	EQ13: How does the New Zealand compliance cost compare internationally?	A feasibility study of whether a robust assessment of New Zealand compliance cost relative to other countries can be made. In particular see if a benchmarking exercise is possible with other jurisdiction(s) also contributing to the cost of the study.	Commissioned: September 2005 Completed: February 2006	$50,000-$100,000
EP11: International compliance cost comparative estimate	EQ13: How does the New Zealand compliance cost compare internationally?	Proceed in the light of results from EP3 above.	Potentially commissioned: March 2006 Completed: March 2007 (or if framework put in place could be an ongoing study)	$0 (if not done) - $300,000
EP12: Regular review of what other jurisdictions are doing	EQ16: What can be learnt from other jurisdictions for improving the system?	DBH staff	Completed: March 2006, 2008, 2010	Low

Evaluation Project (EP)	Evaluation questions	Way of proceeding	Timing	Estimated cost[12]
EP13: Other evaluation information not requiring separate project	EQ17: Are customers satisfied with new building work?	DBH staff	Ongoing	Low
EP14: Review of materials certification system	EX19: Is the materials certification system working effectively?	Expert review of materials certification system based on indicator information, document review and key informant interviews.	Commissioned: March 2007 Completed: December 2007	$30,000-$50,000
EP15: Review of building practitioner licensing system	EX20: How well is the building profession registration system working?	Expert review of building practitioner licensing system based on indicator information, document review and key informant interviews. This could provide a snapshot which could then be compared over time with later one.	Commissioned: March 2010 Completed: December 2010	$30,000-$50,000

VII. Evaluation Management and Implementation Considerations

A. Evaluation Management Structure

Evaluation management structure needs to provide effective governance and management for the ongoing planning, implementation and reporting from the evaluation projects under the new building regulatory regime evaluation strategy. It is recommended that the following be put in place:

- An explicit position of *overall evaluation manager*. This position may be separate or combined with a position responsible for managing DBH monitoring. It requires evaluation management skills and if the manager does not have these skills, steps should be taken to upskill whoever is in this position by having them attend appropriate courses and conferences.
- Access to *evaluation specialist skills* for oversight of the evaluation projects. Evaluation specialist skills are required if DBH is to maintain sufficient oversight of the evaluation projects that will be undertaken under this evaluation strategy. These skills can be obtained by either employing someone in-house or by contracting in advices as and when needed.
- Access to *skilled evaluation practitioners* to undertake the evaluation projects identified in this strategy. The advantages and disadvantages of internal and external evaluation staff are set out in the table below.

Internal evaluation project staff	**External evaluation project staff**
More integration with strategic planning and the rest of DBH	Less integration with strategic planning and the rest of DBH
Potentially less independent in evaluative judgments	Potentially more independent in evaluative judgments
Lower cost	Higher cost
Given the shortage of evaluation skills potentially less	Potentially higher skilled if from a skilled and experienced

skilled	evaluation consultant or organisation
With good knowledge management infrastructure more likely to retain institutional knowledge	More likely to not contribute to institutional knowledge
Potentially distracted by other work priorities within DBH	Less likely to be distracted by other work priorities within DBH
Easier to maintain control of and potentially less evaluation question drift	Harder to maintain control of and potentially more evaluation question drift

- An *evaluation committee* responsible for the evaluation strategy and evaluation projects. There are three functions that such a committee needs to fulfill: 1) being the keeper of the evaluation strategy (this is achieved in practice by them ensuring that the REMLogic structure remains up-to-date and that it drives evaluation planning, implementation and linkages to other aspects of the DBH such as strategic planning and monitoring; 2) oversight of the implementation of evaluation projects; 3) technical and strategic input into evaluation projects. In some cases evaluation projects have both a technical advisory committee providing technical advice on the evaluation and an overall evaluation steering committee playing a governance function, however this requires additional resources for the setting up and running of two committees. It is suggested that in the first instance a single evaluation committee is established to oversee this evaluation strategy. This committee should include: key DBH managers (obviously the evaluation manager and whoever is responsible for DBH monitoring); a senior manager as an evaluation *sponsor* keeping evaluation as a live issue in the highest DBH management level; key external stakeholders who may be involved in the evaluation in various ways (e.g. TAs, BRANZ and the building industry) and one or more evaluation specialists.

B. Knowledge Management

Effective knowledge management is essential for sound evaluation management, it is therefore recommended that the following steps are taken in regard to knowledge management for the evaluation of the new building regulatory regime:

- The REMLogic structure is used as the heart of ongoing evaluation knowledge management and placed on an intranet. All evaluation questions within evaluation projects should be related back to the evaluation questions identified in the REMLogic Evaluation Questions Table. If evaluation questions are changed in the course of designing and implementing specific evaluation projects, the appropriate section of the REMLogic Evaluation Questions Table and the Evaluation Projects List should be changed to reflect this. The purpose of this is to maintain a *living REMLogic Structure* which at any time provides an up-to-date summary of evaluation planning and progress in implementing the evaluation strategy. This ensures integrated planning; allows the Evaluation Committee to get a rapid overview of how evaluation implementation is tracking; and eliminates the need to prepare separate summaries of progress on the evaluation when these are called for from time to time.
- Hyperlinks out beneath the Evaluation Projects List should provide access to all evaluation documentation (including RFPs, evaluation reports etc.).
- The Evaluation Committee maintain an updated *Frequently Asked Questions* about each evaluation project which documents the important decisions made in regard to each evaluation project and that this document is hyperlinked beneath the relevant project in the Evaluation Projects List.
- That as a part of all evaluation projects there is the requirement that an evaluation findings summary be provide in a suitable format to hyperlink beneath the appropriate part of the outcomes hierarchy and also the appropriate project in the Evaluation Project List; doing this will tie evaluation reporting and findings directly back to the outcomes hierarchy. If the outcomes hierarchy is then used for ongoing strategic planning, this approach will encourage a direct feed of evaluation findings back into periodic strategic planning.
- That DBH identify an efficient way of managing this knowledge structure. There are various ways this could be done which do not need to be expensive. One way is to simply use Inspiration[14], the programme in which the outcomes hierarchy was drawn and which allows hyperlinks to documents beneath it. Documents can then be in any suitable format such as Microsoft Word.

[14] Inspiration can be obtained from www.inspiration.com for approximately $150NZ

C. Risk Management

The risks which need to be managed in this evaluation are listed in the table below.

Asking and answering the right evaluation questions.	The REMLogic structure sets out the evaluation questions and the rationale for why they have been selected in the evaluation. This should be subject to peer review in order to mitigate this risk.
Obtaining evaluators with the right skills to undertake the evaluation.	This is an ongoing problem due to the current shortage of skilled evaluators. This risk can be mitigated by advertising RFPs as widely as possible to lists of evaluators, including through the Australasian Evaluation Society. In addition, individual evaluators could be approached as was done in seeking responses to the RFPs for the evaluation planning phase of this project.
Drift in evaluation questions an evaluation project is answering and stakeholders not understanding exactly what evaluation questions are being answered.	This occurs where an evaluation project starts off attempting to ask one evaluation question but progressively drifts away from this question as methodological and practical problems arise. This can result in the evaluation report answering a different question from that which stakeholders initially thought was being answered. The REMLogic approach forces explicit consideration of the feasibility and cost of answering evaluation questions at the start of evaluation planning and this reduces this risk. In addition, where this risk is high in regard to an individual evaluation project, it can be managed by the first stage of the project being a stand alone feasibility study. The completion of this feasibility study provides a decision point as to whether it is sensible to proceed to attempt to answer the specific evaluation question under consideration.
Lack of effective control of evaluations due lack of knowledge of evaluation methodology and to turn over of DBH staff and hence loss of institutional knowledge.	The first issue of knowledge of evaluation methodology can be reduced if there are DBH staff who are knowledgeable about evaluation methodology. Alternatively, or in addition to this, an evaluation specialist can be employed to be on evaluation advisory committees. The second issue of staff turnover creates a major problem in maintaining control of evaluation projects. This problem can contribute to evaluation drift as discussed above, repetitive relitigation as to why certain evaluation questions are not being asked, and in some cases to evaluators being criticised for simply implementing design decisions which were made by earlier iterations of the controlling evaluation committee. This risk can be reduced by maintaining a Frequently Asked Questions paper which is updated after each evaluation committee meeting and which progressively documents the major decisions which have been made in regard to the evaluation design. This document should be hyperlinked behind the REMLogic structure (from the Evaluation Projects table). The employment of an outside evaluation specialist who continues to attend evaluation advisory committees while staff turnover means

	departmental staff change, also provides much more continuity to discussions in such committees.
Lack of integration of monitoring and evaluation.	The REMLogic approach, if it continues to be consistently applied in the future by the DBH, should prevent this risk from occurring as it explicitly links monitoring and evaluation into an integrated strategy.
Disconnect between evaluation projects and ongoing strategic planning.	The REMLogic approach, if institutionalised within DBH, can ensure that there is a connection between evaluation planning and findings and ongoing strategic planning by DBH. Institutionalisation can be achieved by integrating the outcomes hierarchy developed as part of the REMLogic structure with DBH internal strategic planning processes and their related frameworks and diagrams. If annual strategic planning is based around discussing how to better achieve the intermediate outcomes set out in the a REMLogic type of outcomes hierarchy, then this can be used to ensure that evaluation findings (already linked in REMLogic to specific intermediate outcomes) are linked back directly into strategic planning discussions. In addition, forward evaluation planning should take place at the same time as strategic planning and this is facilitated by using the REMLogic approach.[15]

[15] Linking strategic planning to evaluation planning is discussed further in Duignan, P. (2004). *Linking Research and Evaluation Plans to an Organisation's SOI.* http://www.strategicevaluation.info/se/documents/120pdff.html

VIII. APPENDIX ONE: EVALUATION DESIGNS AND METHODS FOR CONSIDERATION WHEN USING THE REMLOGIC METHODOLOGY FOR PLANNING EVALUATION ACTIVITY

Dr Paul Duignan

Parker Duignan Ltd

paul@parkerduignan.com

www.strategicevaluation.info

V1-3 June 2005

There is a wide range of evaluation methods and designs available to collect evaluation evidence. Evaluation methods are specific methods used to collect information in an evaluation project, for example, *surveys* or *interviews*. Evaluation designs are the way in which an evaluation project is structured, for example, as an *experiment* or a *case study*[16]. This paper sets out information on methods and designs. If using information from this paper please cite it to the reference given below.

A summary of the most common evaluation methods is as follows:

[16] This discussion is drawn from Duignan, P. (2001) *Introduction to Strategic Evaluation: Evaluation Approaches, Purposes, Methods and Designs.* www.strategicevaluation.info/se/documents/104f.html. If using this material please acknowledge source.

Evaluation Methods

- *Document thematic analysis* (sometimes referred to as *desk reviews*). Analysis of documents in order to extract themes in regard to the topic being evaluated. This can include all types of documents such as formal reports, minutes of meetings, memos and print media reports. Where the documents are research reports this is referred to as a literature review. The analysis can range from an extraction of general themes to a tight and specific detailed analysis. If qualitative ratings are used, an element of subjective judgment is inevitable but the risks can be reduced by documenting as explicitly as possible, in advance, what the reviewer would expect to see to warrant specific ratings and by using more than one reviewer.

- *Written surveys.* Surveys can be undertaken with groups of stakeholders who are the users of, or knowledgeable about, aspects of the activity being evaluated. In such surveys, respondents can be questioned on how they view both the quality and the results of the activity. The advantage of respondents' judgments of this type is that they can provide responses which combine multiple factors into an overall summary evaluative judgment. The disadvantage is that such respondents may not be fully knowledgeable or may not, because of their close involvement in the activity being evaluated, be able to provide a sufficiently disinterested view on the worth of the activity. Written surveys also have the disadvantage of potentially low response rates.

- *Telephone or face-to-face interviews.* Stakeholders can also be interviewed either by telephone or face-to-face. The issue of respondents' judgments discussed above in regard to written surveys also applies to telephone and face to face interviews. However, both telephone and face-to-face interviews have the great advantage of allowing for interaction; such interaction lets the interviewer pursue details of points respondents make both in the interview itself and also to seek subsequent interviewees' perspectives. However, undertaking interviews is more expensive than written surveys, and therefore fewer respondents will be able to be contacted for a similar cost. In most settings, telephone or face to face interviews achieve a higher response rate than written surveys.

- *Stakeholder focus groups and group interviews.* Group interviews are used as a cost-effective way of interviewing more than one stakeholder at the same time. There is nonetheless a trade-off between group interviews and individual interviews in terms of the amount of time each stakeholder has to respond. Focus groups are group interviews where there is an interest in having participants discuss an issue *amongst themselves* in order to provide a richer perspective on the topic than can be obtained by either individual interviews or group interviews where respondents spend all the time just answering the questions as individuals.

- *Statistical and econometric analysis.* Statistical and econometric analysis is any quantitative data analysis used to analyse aspects of the activity being evaluated. The data being analysed can be routine statistics collected by an organisation itself, that collected by other organisations for other purposes, or the results of data collection within specific evaluation projects.

- *Case studies.* Case studies look in detail at what happened in a specific instance. They use a variety of methods and more than one case study can be linked into a multiple case study. They enable conclusions to be drawn about how the activity being evaluated occurred in a particular instance.[17]

Evaluation Designs

Evaluation designs are the way in which evaluation methods are used and in some cases combined in order to draw out evaluation conclusions. Evaluation designs can incorporate any of the methods listed above. There are various ways of

[17] Case studies can also be viewed as a type of evaluation design. However for the purposes of this paper they are considered as a method which can be employed in one of the designs listed in the design section

classifying evaluation designs. The following method attempts to break evaluation designs into a small set of headings which can be used when assessing the feasibility of answering evaluation questions.

Formative evaluation designs – evaluation activity to optimise programme implementation

- *Programme-staff self-evaluation formative evaluation design.* Programme staff use various methods to step out of the roll of running the programme and critique what they are doing and optimise its implementation.

- *External formative evaluation design.* A formative evaluation individual or team works alongside a programme as a *critical friend* critiquing and assisting the programme to optimise its implementation.

- *Peer cross programme formative evaluation design.* Representatives of a number of programmes get together to peer-evaluate the implementation of the programme to optimise its implementation.

Process evaluation designs – evaluation activity to describe the history, course or context of a programme

- *Programme description and documentation process evaluation design.* Case study methods are used to provide a rich description of what the programme consisted of.

Outcome evaluation designs[18] – evaluation activity to measure the positive and negative intended and unintended consequences of a programme

[18] More detail on outcome evaluation designs can be found in Shadish, W., Cook, T. & D. Campbell (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

- *Experimental outcome evaluation designs.* Where a comparison is set up between a group which receives the intervention and a group (ideally randomly selected from the same pool) which does not. If not applicable at the higher-levels in an outcomes hierarchy (as is often the case in multifaceted multi-outcome social programmes) they may be applicable at lower levels. These designs should be considered in any instance where one group can be treated in a different way from another group and the resultant outcomes measured.

- *Regression discontinuity outcome evaluation designs.* Where entities (e.g. people, organizations etc.) can be rated in terms of an outcome variable and where it is ethical and practical to only provide an intervention to a sub-set of entities, it is theoretically possible to use a regression discontinuity design. This is where the intervention is only provided to those entities below a cut-off point on the outcome variable. Once sufficient time has passed for the treatment to have had an effect, the new data on the outcome variable is examined for each entity and if the novel approach has had an effect there should be a clear discontinuity at the cut-off point on a graph of outcomes for all entities.

- *Interrupted time series outcome evaluation designs.* Tracking a measure of an outcome a large number of times (say 30) and then looking to see if there is a clear change at the point in time when the intervention was introduced. These require a clearly time of introduction of the intervention and may or may not be interpretable depending on the specific situation.

- *Constructed matched comparison group outcome evaluation design.* Attempting to locate a group which is matched to the intervention group on all important variables apart from not receiving the intervention.

- *Causal identification and alternative explanation elimination outcome evaluation design*. These designs attempt to identify all of the possible causes for an outcome, collect evidence for and against each of them and through a process of elimination identify one or more causes for an outcome. This is in contrast to some pseudo-outcome studies which simply collect evidence consistent with attribution of changes in outcomes to one treatment without examining whether the evidence is not also consistent with attributing changes in outcomes to alternative factors.

- *Expert connoisseurship outcome evaluation design.* Where one or more subject experts are asked to analyse a situation in a way that makes sense to them and to assess whether on balance they accept the hypothesis that the intervention may have caused the outcome.[19]

- *Stakeholder judgement design.* Where a number of key informant from stakeholder groups are asked to analyses a situation in a way that makes sense to them and to assess whether on balance they accept the hypothesis that the intervention may have caused the outcome.[19]

[19] The last two designs would not usually be expected to establish causality as robustly as the other listed designs. However these designs are frequently used and deserve a place in a full typology of outcome evaluation designs; in particular circumstances they are feasible, affordable and accepted by stakeholders as better than having no high-level outcome attribution information. Even though they are often more feasible and affordable than the other designs, decision-makers have to consider on a case by case basis whether these designs can actually provide any coherent information about attribution or whether they will just end up being examples of pseudo-outcomes studies which do not contribute any sound information about attribution.