Report No.1874
April 2011

Methods for Developing Sustainability Indicator Systems for Freshwater and Estuarine Receiving Bodies of Urban Stormwater

RESEARCH BASED SOLUTIONS

Methods for Developing Sustainability Indicator Systems for Freshwater and Estuarine Receiving Bodies of Urban Stormwater

Chris Batstone[1]

James Baines[2]

Eric Goodwin[1]

Bronwyn Morgan[2]

Toni Canard[3]

1: Cawthron Institute; 2: Taylor Baines & Associates; 3: CSIRO Cleveland, Queensland Australia

Prepared for

NIWA

Cawthron Institute
98 Halifax Street East, Private Bag 2
Nelson, New Zealand
Ph. +64 3 548 2319
Fax. + 64 3 546 9464
www.cawthron.org.nz

Reviewed by: Jim Sinner

Approved for release by: Rowan Strickland

Issue date: 20 April 2011

Recommended citation:
Batstone CJ, Baines JM, Goodwin E, Morgan B, Canard T 2010. Methods for Developing Sustainability Indicator Systems for Freshwater and Estuarine Receiving Bodies of Urban Storm Water. Prepared for NIWA. Cawthron Report No.1874. 55 p. plus appendices.

EXECUTIVE SUMMARY

This report describes options for developing sustainability indicator systems. The possible methods are numerous, and range in spatial scale from global to national to local. The applications for such indicator systems vary from whole system scope to issue focus. This report focuses on the development of a sustainability indicator system for a specific application: understanding the sustainability implications of alternative urban development strategies for two kinds of receiving bodies for urban run-off stormwater: streams and estuaries. The indicator system is intended to form part of a Decision Support (DSS) tool that utilises a Bayesian Belief Network in which the precursor variables of the composite indicator are derived and configured.

This work integrates governance considerations, indicator theory, and existing systems to derive recommendations for a methodology. The technical basis for the method was derived from the work of the Statistics Directorate of the OECD (Organisation for Economic Co-operation and Development) and various journal papers and manuals that describe multi-criteria methods. The options for technical approaches fall into two broad categories: expert and stakeholder driven. The differences lie principally in how the data for the indicators is derived and processed. In the composite indicator (CI) approach, a highly quantitative approach favoured by the OECD, statisticians determine the relative importance of the data that are integrated to form the composite indicator according to statistical criteria. In contrast, in a Multi-Criteria Analysis (MCA) approach stakeholders determine how items are weighted according to qualitative criteria.

A review of literature on governance, and specifically on integrated sustainability, highlights the importance of clarifying the sustainability definition within which the DSS is operating. In broad terms sustainability concepts are expressed as “weak” or “strong”. Weak sustainability recognises transfer of capital between natural and financial forms. Strong sustainability denies the transfer of capital from natural to financial forms and emphasises the need to at least preserve the current environmental endowment without further degradation. The Resource Management Act (1991) is a standards-based approach to environmental management that suggests the appropriate sustainability objectives are more aligned with strong sustainability than weak sustainability. This in turn implies a target or reference point approach in the construction of composite indicators. The issue of the appropriate sustainability objective for the DSS is one that should be resolved in stakeholder consultations.

A key lesson from the governance review is the need for early and frequent stakeholder consultation. The roles and interaction between stakeholders and decision makers in the process in which the DSS operates should be clarified. This has implications for the configuration of the DSS in terms of user definition and stakeholder accessibility. The potential for the tool to support adaptive governance and social learning processes should not be underestimated and should be evaluated.

A further outcome of the governance review lies in the extent to which the four “pillars” of sustainability are integrated. While recognising the need to assess the economic, social, environmental and cultural implications of alternative urban development strategies, contemporary best practice is moving away from separate assessment towards combining information from those domains into integrated assessments that present information relevant to targets and reference points that reflect "limits".

The most comparable existing system to that under consideration here is a web-based application of MCA called "DayWater". DayWater is a large, complex, and ambitious European Union (EU) project that built a DSS accessible by stakeholders as opposed to having decision makers as the primary focus. The limitations of the DayWater DSS lie in an absence of a catchment-based spatial focus, and the lack of an inter-temporal aspect in which changing environmental responses to urban development options are depicted for input to adaptive governance processes.

Beyond incorporation of explicit spatial and inter-temporal aspects in the DSS, other opportunities to build on existing applications lie in taking a whole system approach to assessment, and the inclusion of resilience concepts to the indicator matrix. In a whole system approach, consideration of economic, environmental, social and cultural effects of alternative urban development strategies for urban run-off stormwater extends beyond the specific freshwater and estuarine receiving bodies to include terrestrial and air column effects. This expansion of scale allows a wider scope of trade-offs inherent in policy development to be included in the DSS.

In the context of environmental management, "resilience" is an emergent concept from the field of socio-ecology that refers to the capacity of complex adaptive systems to absorb shocks while retaining structure and function. Complex adaptive systems are characterised by high levels of uncertainty, thresholds, and rapid, sometimes irreversible, change from one stable state to another. A resilient ecosystem can withstand shocks and rebuild itself. Resilient social systems have the added capacity for humans to anticipate and to plan for the future. This branch of environmental science advocates the development of resilience indicators, for example identifying thresholds of change and resilience capacity, as a key area for development in environmental management.

Recommendations for a methodology combine considerations from both CI and MCA approaches. To meet stakeholder consultation requirements, a mediated modelling approach using a Bayesian Belief Network model and an Analytical Hierarchy Process (AHP) is favoured to establish data inputs and relative weights between sub-indicators for the indicator system.

Specific technical recommendations for the composite sustainability indicator method include:

- Use of the Bayesian Belief Network (BBN) approach to create a conceptual map of the issue and for generating the pre-cursor variables and the associated probabilities on which composite indicators are based,
- Investigation of two approaches to normalisation of precursor variables: a rescaling approach and distance to reference values,
- The use of Analytical Hierarchy Process (AHP) to generate weights for the pre-cursor variables,
- The investigation of two approaches to aggregation of pre-cursor variables into a composite indicator: geometric aggregation and choice modelling.

Development of the sustainability indicator system could proceed in two phases.

In the first phase the scope of the indicators would be limited to the estuarine and freshwater receiving bodies, with the DSS tool focused on providing planners and technical experts with depictions of the effects of alternative urban development scenarios.

In the second phase the tool would be expanded to encompass the entire social-ecological system with capacity for use in adaptive governance and social learning processes. This implies the development of indicators for resilience in the social-ecological context of the UPSW (Urban Planning for Sustainable Waterbodies) project. The resilience concept provides a framework with a reasonable pedigree to move from system definition focused on the biophysical receiving water bodies to a whole system focus which spans the entire social-ecological system of human urban development and associated catchment and downstream effects.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION

This report presents recommendations for developing a sustainability indicator system. The possible methods are numerous, and range in spatial scale from global to national, to local. The applications for such indicator systems are also varied, from whole-system scope to specific-issue focus. This report focuses on the development of a composite sustainability indicator system for a specific application: understanding the sustainability implications of alternative urban development strategies for two kinds of receiving bodies for urban run-off stormwater: streams and estuaries. The sustainability indicator system is intended to form part of a Decision Support (DSS) tool and Bayesian Belief Network in which the precursor sub-indicator variables of the indicator are derived and configured.

This work integrates the governance context, indicator theory, and existing systems to derive recommendations for a methodology. The options for technical approaches fall into two broad categories: expert and stakeholder driven. The differences lie principally in how the data for the indicators are derived and processed. In the highly quantitative technical approach data are drawn from surveys or other similar sources. Statisticians determine the relative importance of the data that are integrated to form the composite indicator according to statistical criteria. In contrast, in a Multi-Criteria Analysis (MCA) approach stakeholders determine how items are weighted according to qualitative criteria.

The report is configured as follows. Section 2 reviews the literature on integrated assessment and governance. Section 3 describes specific approaches found in the sustainability indicator technical domain. Two key approaches are identified, with a number of alternative and complementary techniques reviewed: the composite indicators (CI) approach and Multi-criteria analysis (MCA). A SWOT analysis is undertaken of each. Section 4 recommends a methodology for developing a composite sustainability indicator system that effectively discriminates between the sustainability implications of alternative urban development options. Section 5 presents conclusions. Appendix 1 contains a bibliography of the papers reviewed.

Appendix 2 contains the full text of the report prepared by Taylor Baines & Associates entitled, “Sustainability Assessment - Governance and Management Issues: Literature Report.” Appendix 3 contains the “Handbook on Constructing Composite Indicators: Methodology and User Guide” compiled by the Statistics Directorate of the OECD.

2. SUSTAINABILITY ASSESSMENT - GOVERNANCE AND MANAGEMENT ISSUES

This section of the report describes desirable features of a sustainability indicator system from the perspective of governance and management, based on a literature review by Taylor Baines & Associates.

Section 2.1 summarises the relevant literature and positions the UPSW research in the context of governance and management issues associated with sustainability assessment. Section 2.2 identifies the requirements for implementing an integrated sustainability indicator system (a DSS based on Multi-Criteria Assessment tool) and makes recommendations for the UPSW process.

The full version of the Taylor Baines & Associates report, from which this section of the report has been drawn, is located in Appendix 2.

2.1. Requirements for effective sustainability assessment

The literature canvassed in this review has drawn from experience in North America, Europe and Australasia. Considerable cross-fertilisation of ideas and initiatives is evident from the bibliographies, although clustering into distinctly separate citation sequences is also evident. The literature search encountered a number of emerging and over-lapping fields of enquiry, experimentation and practice. For simplicity, the term Sustainability Assessment will be used hereafter to include such variants as Sustainability Appraisal, Integrated Assessment or Integrated Sustainability Assessment.

Formal applications of Sustainability Assessment encountered so far are predominantly at the strategic level. That is to say, they are generally aimed at public policy decisions or strategic public agency decisions. Analytical frameworks for Sustainability Assessment are still generally at the developmental or trial stage. However, these same characteristics - strategic level and developmental stage - are precisely those which apply to the research context for this literature review. When this field was reviewed in 2006, Brinsmead (2005) was the only literature which identified at a fundamental conceptual level what integrative functions are necessary for the practice of integrated assessment. This review update has been unable to re-contact Brinsmead; nor has it found any further elaborations or extensions of his work; nor has it found any other comparable analysis or first principles-based review of the practice of 'integration' in assessment activities. However various authors (Gibson 2006; Sadler 2008; Videira *et al.* 2010) have attempted to make progress in articulating fundamental principles of good Sustainability Assessment practice.

Two main themes have emerged from this review of integrated assessment. The first theme is the increasing recent focus on Sustainability Assessment as a structured assessment methodology, as distinct from the earlier more general theme of integration in assessment. Within this methodological focus, sub-themes include:

- Perspectives on sustainability,
- Identifying objectives for assessment and decision making,
- Clarifying trade-off rules,
- Adaptive processes, and
- The relationship between those involved in assessing and those involved in decision making.

The second main theme is stakeholder engagement as a specific focus of practice, exhibiting a maturing interest and more sophisticated discourse about rationales and approaches to stakeholder engagement in Sustainability Assessment processes. Within this focus, the sub-themes are:

- Greater differentiation and specification of stakeholder roles and functions, and
- The phenomenon of social learning.

2.1.1. Perspective on sustainability - characterising the current position

Several writers with longstanding involvement in the development of assessment tools preface their discussion of practical Sustainability Assessment methodologies by conceptualising the pursuit of sustainable development primarily in terms of remedying what can clearly be seen already to be unsustainable and at risk, rather than aiming for ill-defined and difficult to agree aspirations.

Their framing of the sustainability challenges is similar to the framing of the UPSW concept, with its identification of impacts on receiving waters (environmental capital) of current stormwater management practices and the reflection of these impacts in ecological, social and financial values. In short, decision making aids such as Sustainability Assessment should incorporate explicit consideration of limits and targets as they relate to the management of capital stocks, and most likely need to recognise that some of these limits may already have been transgressed in particular locations or may be at risk of being transgressed (Gibson, 2006; Sadler 2008).

2.1.2. Identifying objectives to be applied in the assessment - sustainability concept definition

The step change from earlier versions of assessment practice is the shift from relative assessments which focus on assessing whether or not proposals or policies result in outcomes which are generally in the desired direction ("direction to target" assessments) to absolute or comparative assessments which focus on assessing how far a particular proposal or policy will steer outcomes towards desired targets, *i.e.* "distance to target" assessments (Pope *et al*. 2004).

For such assessment procedures, objectives have to be defined in a manner which renders them amenable to quantification or qualitative comparison. With regard to practical methodologies for assessment and decision making, the central debate here is between those who adopt a 4-pillar approach (Sadler 2008) to defining objectives and associated criteria and indicators, and those (Kemp *et al.* 2005; Gibson 2006) who attempt to define sustainability objectives in terms of the observed 'requirements for improvement' rather than the established categories of wellbeing.

In the former approach, integration is achieved when economic, environmental and social factors are addressed simultaneously, and evaluated against a sustainability framework derived from international or national policy or strategies. The UPSW research could be interpreted as moving beyond the pillar approach by focusing on "requirements for improvement" in terms that are recognisable to members of the public - underfoot conditions, water clarity and ecological health. It also encapsulates elements of the 4-pillar approach with its simultaneous consideration of environmental, social and economic factors, and may incorporate the notion of upper and lower thresholds (sustainability targets and limits), depending on how the low/medium/high ranges are specified.

2.1.3. Clarifying trade-off rules - sustainability concept application

A common theme in literature on sustainable development is the question of balance in the simultaneous pursuit of multiple objectives (Brinsmead 2005; Kemp 2005; Gibson 2006; Sadler 2008; Listowski *et al*. 2009; Videira *et al*. 2010). It is difficult to avoid the concept of trade-offs, and the related questions of how trade-offs should be conceived in a sustainability context and under what circumstances they should be considered. Propositions for trade-off rules tend to adopt a tiered approach to specifying principles, subject to a general requirement not to compromise the achievement of net sustainability gain. They also emphasise requirements for transparency and an obligation on trade-off proponents to provide explicit justification.

Good practice requires that any trade-off assumptions and mechanisms that are <u>implicit</u> within an indicator development process and embodied in the DSS design should be made <u>explicit</u> and discussed with the decision makers (Gibson 2006; Sadler 2008).

2.1.4. Adaptive process design in Sustainability Assessment

Two aspects of 'adaptiveness' were encountered in the literature. The most common interpretation of 'adaptiveness' relates to iterative assessment processes which have evaluation and learning procedures built into them to enable adjustments in decisions over time (Brinsmead 2005; Kemp 2005; Gibson 2006a; Sadler 2008; Tuinstra *et al*. 2008; Videira *et al*. 2010).

The other interpretation of 'adaptiveness' refers to the need to customize assessment processes to the local context (Gibson 2006; Deutsch and Metelka 2008; Sadler 2008; Thevenot 2008). In the context of DSS development, the following considerations should be given careful attention:

- the intended audience,
- the functionality of the tool in terms of input and output interfaces,
- the nature of the operating software, and,
- the choice of operational location – internet versus desktop should be considered.

It follows that the development of a DSS is an opportunity to enhance the Sustainability Assessment process by providing a tool that is customized to the context, and by deployment in iterative assessment.

2.1.5. The relationship between assessment and decision making

The critical issue here for good practice in Sustainability Assessment is the relationship between the activities of assessment and decision making. Are the decision makers themselves involved directly in the assessment process and thereby intimately familiar with the mix of information and value judgments which have informed the results? Gibson (2006) and Sadler (2008) take a firm view that the direct involvement of decision makers in the assessment activities is a central element of integration in the context of Sustainability Assessment.

2.1.6. Greater differentiation and specification of stakeholder roles and functions in integrated assessment

The literature reveals considerable evidence of a progressive shift from expert, managerial processes towards deliberative democracy, and provides a variety of explanations for this. This range of justifications points clearly to the potential scope for stakeholder involvement throughout every stage of an integrated assessment process, from issue framing and objective setting, through deliberation and prioritising to decision making. When addressing the topic of stakeholder engagement, an inclusive interpretation of 'stakeholder' is advised (Kemp 2005; Lai *et al*.2008; Tuinstra *et al*.2008; Videira *et al*. 2010).

This review confirms the importance of stakeholder involvement to the UPSW research programmed anticipated by NIWA/Cawthron. How the research team has already engaged stakeholders or intends to address questions of stakeholder engagement is clearly important, for all the reasons outlined.

The literature makes it clear that initiatives on stakeholder engagement cannot be subject to rigid guidelines but need to be adaptive to the task at hand. Such initiatives inevitably involve an element of experimentation, informed by the accumulating practical experience of facilitated participation processes (that exists within the research and professional community) and a commitment to evaluate effectiveness periodically.

2.1.7. Social learning

The literature emphasizes the importance of social learning to the contemporary practice of Sustainability Assessment. In the context of integrated sustainability assessment the social learning concept is explained in terms of,

> *"Social learning involves participants learning alongside each other about issues, but also about each other in respect of different perspectives on the issues and the basis of these different perspectives, so providing opportunity for mutual understanding and mutual influence. Social learning may enhance the capacities of those involved in a participatory assessment process (especially those who hold actual or potential agency in respect of problem causation or solution) to change the original institutional conditions and contexts, which 'frame' the issues and which potentially limit (unnecessarily) the possibilities to find solutions."*
>
> Tuinstra *et al*. 2008:131

Certain research tools, such as MCA and mediated modeling, can be helpful in structuring social learning processes (Proctor and Drechsler 2006; Hermans *et al.* 2007; Salgado 2009).

Of particular interest here are the references made by a number of authors to the use of Multi-Criteria Assessment/Evaluation (or similar descriptors) as tools to facilitate social learning. For example, Proctor and Drechsler (2006:172-3) suggest -

> *"MCE has the advantage of being able to provide a framework for complex decision-making problems that allows the problem to be broken down into workable units and to be structured in such a way that the complexities of the problem can be unraveled. This is done essentially through the process of identifying options, criteria, and preferences. Applying an MCE in a heuristic way enables the MCE to aid in the learning process concerning complex issues."*

Similarly, Hermans *et al.* (2007:543), referring to the use of a Multi-Criteria Decision Aid to assist a multi-stakeholder group to evaluate proposed river management options, pointed out that "the process was used primarily as a social learning tool." They added that the structure of the process would be similar whether the outcomes were being used primarily to facilitate discussion and group learning or to achieve consensus.

Tuinstra *et al*. (2008:132) provide some guidance on conditions affecting social learning and on approaches to evaluating the extent of social learning, which are of potential relevance in the research context of this review. Drawing on the evaluation of three integrated assessment processes "where the attempt has been made to institutionalize learning procedures into the process design", they report that participants identified the following key conditions that affect social learning -

- the motivation and skills of leaders and facilitators;
- clarity about the role and purpose of stakeholder involvement;
- the internal structure (of the group) and the room given to democratic debate about regulatory institutions;
- the structural capacities for interaction between social networks; and,
- the influence exerted by the existing culture on the framing and definition of the issues at stake.

In terms of integrated sustainability assessment practice, achieving social learning outcomes amongst participating stakeholders and the NIWA/Cawthron research team will be an important test of the effectiveness of any resulting DSS.

2.2. Requirements for implementing an integrated sustainability indicator system (an DSS based on Multi-Criteria Assessment tool)

The discussion in Section 2.1.7 above regarding the capacity of certain decision processes to generate social learning outcomes motivates review of a number of Multi-Criteria Assessment (MCA) cases that have been applied in the context of Sustainability Assessment and related decision making. Most of these are cases of applying a general methodology to a particular sustainability challenge. Of the cases reviewed, the DayWater Project (Thevenot 2008) is unique in attempting to produce a computer-mediated DSS tool for remote access by users. All the other cases involve the use of MCA methodologies in interactive group processes, although some of these involved use of computer-based tools mediated by the researchers.

This section examines the expressed strengths and weaknesses of such MCA processes to see if they provide additional insights to those already gleaned from the broader Sustainability Assessment literature.

2.2.1. Strengths, weaknesses and process issues encountered in multi-criteria applications

This sub-section of the report considers multi-criteria methods from a governance/management perspective. Multi-criteria methods are explored in greater detail in Section 3.4 where they are contrasted with compiste index approaches to indicator development. Strengths and weaknesses are here identified from the perspective of experts and stakeholder participants. The DayWater Project was reviewed from the reported experiences of the MCA tool developer and a prospective end user to see what additional guidance could be gleaned for the NIWA/Cawthron DSS development work.

Strengths from an expert perspective

The strengths of MCA processes from an expert perspective are associated with their contribution to clarifying logic and the fact that they allow values to be incorporated into assessments in an explicit manner.

For clarifying logic, it is claimed that,

- MCA processes allow a complex problem to be broken down into workable units, where options, criteria and preferences are made explicit,
- Simpler MCA methods can encourage transparency of logic (valued by stakeholders) and provide a useful structure for communicating decisions,
- The MCA framework assists in structuring preparations and deliberation processes; in making information transparent (alternatives, criteria, impact scores, weights) and pointed to a clear structure for each participatory process.

For getting values incorporated, it is claimed that,

- MCA processes result in a good combination of agreed facts and social values,
- Citizens are involved to make comparative values judgments in a long-term perspective,
- Citizen juries can be used to aggregate multiple individual preference weights through deliberation to achieve consensus.

Strengths from a stakeholder/participant perspective

The strengths of MCA processes from stakeholder/participant perspective are associated with the perceived benefits of structured and open deliberations, the perceived legitimacy of their participation, and the opportunity to learn and understand complex issues better. The benefits of structured and open deliberations experienced are described as,

- Giving citizens increased understanding of different points of view,
- Enabling the group to learn and ‘move forward’,
- Encouraging participants to focus more on the preferences and weightings for the criteria than on the final outcome, and
- Framing of citizen deliberations to favor social values -- the format did not preclude voicing of individual interests, but did tend to de-legitimize them.

The perceived legitimacy of participation is reflected by,

- The involvement of interest groups (clarifying different interests) and citizens groups (comparison of interests) being seen as complementary,
- The method of interest group involvement focusing on explanations of their perspectives rather than lobbying for specific solutions.

Enhanced learning is provided for,

- Once the MCA tool is familiar to users, the ease and speed of use allows a degree of ‘experimenting’ with weights to test concerns and options, or when applied in a heuristic (iterative with feedback) way, the MCA tool aids jury learning about complex issues, and

- By the use of maps and visual illustrations of alternatives to aid transparency.

Weaknesses from an expert perspective
The weaknesses of MCA processes from an expert perspective are associated with data problems, methodological challenges, process facilitation difficulties and resourcing needs for the assessment.

Data problems may be associated with:

- Difficulty in finding quantitative and/or accurate information on many of the criteria identified for assessment,
- Difficulties with identifying or agreeing scale descriptors for impacts.

Methodological challenges arise,

- When it is not clear whether the method has dealt appropriately with trade-offs or compromises involving compensation of one factor loss by another factor gain (see Recommendation 1 in Section 2.3),
- Over the assumption that preferences for different criteria are assumed to be independent of each other,
- Through a double-counting problem when chosen criteria are either redundant or non-exhaustive.

Interactions between analyst and decision makers can become difficult if the analyst is taking on dual roles of science expert and process facilitator (see Recommendation 6 in Section 2.3). The resourcing implications which arise if many assessment criteria and/or stakeholders are involved must be taken into account. Stakeholder engagement processes need to be 'fit for purpose' within the constraints of the resourcing available.

Weaknesses from a stakeholder/participant perspective
The weaknesses reported from a stakeholder/participant perspective appear mainly to be associated with excess complexity and an absence of effective social learning. They include,

- The need to balance complexity/simplicity with cognitive capacity; complex MCA methods can be perceived by non-experts as "black box" approaches; too many objectives/criteria can overload individual's thinking and analysis,
- Citizens can become overwhelmed by expert contributions in some situations - distracting them from their long-term focus,
- The differences between scientific knowledge and practical knowledge and the different ways of thinking about a real/abstract situation or problem,
- Having experts/interest group representatives select criteria risks missing some criteria considered important by individual citizens,

- Where citizens juries have not been involved in the task of structuring the decision making.

2.3. Recommendations: extension to the UPSW Research

The review reported in this section of the report leads to the following recommendations for the UPSW development of a sustainability indicator system.

1) Attention to trade-offs

Good practice will require that any trade-off assumptions and mechanisms that may be implicit within the choice experiments and associated modeling should be made explicit and discussed with the decision makers to ensure that they are aware of the nature of the trade-offs associated with any choice/decision (Section 2.1.3).

2) Dialogue throughout the development process

Building in both elements of adaptiveness to the DSS will require close attention to stakeholder/researcher dialogue. In particular, building in evaluation and learning procedures to the methodology to enable adjustments in user decisions over time will require such dialogue to take place (be managed) at various stages of DSS development, not just at the beginning and end of the development sequence (Section 2.1.5).

3) Stakeholder considerations

A selection of intended future users/decision makers should be included amongst the wider group stakeholders to assessments involved directly and throughout the development of the DSS. It should not however be assumed that they are the only stakeholders or indeed the most important stakeholders (Section 2.1.6).

4) Timing of stakeholder engagement

If stakeholder engagement has not already been initiated, this should happen before the end of the ‘proof of concept’ stage with the expectation of periodic involvement throughout each stage of DSS development (Section 2.1.6).

5) Collaboration with other science programmes

NIWA/Cawthron should consider initiating a science forum devoted to stakeholder engagement issues, with a view to providing a regular forum for exchange of ideas and experience amongst scientists with these common interests.

6) Expert facilitation

In order to maximize social learning during the research programme, the NIWA/Cawthron research team should consider the use of expert facilitators to plan and facilitate periodic group work sessions involving stakeholders.

7) Testing for social learning outcomes

As a means of monitoring progress toward an effective DSS, the NIWA/Cawthron team should consider monitoring social learning outcomes amongst participating scientists and stakeholders at intervals during the development process.

3. CONSTRUCTING SUSTAINABILITY INDICATORS

Two existing systems have been identified as a basis for comparison from which the UPSW project can derive insights.

The first, "Hydropolis" from the DayWater project is reviewed in Section 3.1. The second, the triple bottom line (TBL) process, is reviewed in Section 3.2.

A number of approaches have been identified in the literature for the construction of sustainability indicators. Two have been selected for their representation of key contrasting issues. First, in Section 3.3 the work of the Statistics Directorate of the OECD (Nardo *et al.* 2005), and second, in Section 3.4 Multi-Criteria Analysis (MCA).

3.1. Hydropolis: a Decision Support System

The DayWater Project has been a major international research and development initiative attempting to develop an Adaptive Decision Support System (DSS) for urban stormwater source control, an objective similar to that of the NIWA/Cawthron research team. Of the cases reviewed, the DayWater Project is unique in attempting to produce a computer-mediated DSS tool for remote access by users. All the other cases involve the use of MCA methodologies in interactive group processes, although some of these involved use of computer-based tools mediated by the researchers.

Whilst the NIWA/Cawthron concept is not identical to any other approach encountered in this review, the intention to develop a DSS means that lessons from the DayWater experience are likely to be of interest.

The compendium of papers describing many aspects of the DayWater Project over its progress from 2002 to 2008 is available to the NIWA/Cawthron research team. Its details will not be reproduced here.

At the level of programme design it is particularly useful to have had access to papers written from several perspectives, notably the scientific/technical perspective and end-user perspective. The scientific/technical perspective addresses, amongst other aspects, the design of the overall DSS architecture (Deutsch and Metelka 2008) and the specific MCA tool (Ellis *et al.* 2008), in this case referred to as the DayWater Multi-Criteria Comparator (MCC). The end-user perspective (Beyeler 2008) describes an end user's account of their involvement in and experience of the prototype development.

Some contextual detail is helpful. The DayWater Project was an immense and ambitious undertaking. The Project involved a consortium of ten scientific partners, from the Arctic circle (Sweden) to the Mediterranean Sea (Greece). Thevenot (2008) describes the consortium as "dominated by engineers and urban hydrologists" but also with "experts in socio-economics and information technology". There have also been 14 core end-users, with the intention to extend to a much wider group of end users for testing components of the DSS. Core end users are predominantly state, city or county water agencies, indicating a narrow conception of stakeholders. For geographic and linguistic reasons, local DayWater partners have been responsible for interacting with nearby core end users. Although the extended end user groups have been involved in a series of eight regional conferences, it appears that they have yet to become involved in actual testing of DSS components. Thevenot (2008) reports that end-user involvement was one of the most difficult tasks in the DayWater project, constrained by time demands, language barriers, and the level of mutual understanding achieved. Collecting and collating end-user needs in order to agree the functionality of the DSS was much more demanding than expected, taking 24 months instead of the expected 12 months. Clearly, project coordination was problematic, facing difficulties of multi-disciplinary and multi-national dimensions –

> *"even if scientific partners are working in the same field, there is a common language to be found for the project. ....Each partner developed their components mostly independent of the other components and their mutual requirements. .... The analysis of the decision making procedure in urban storm water management projects is a good example of this kind of intercultural and interdisciplinary misunderstanding. Consulting engineers generally conceive the decision making procedure in a different way than researchers and south European countries do not consider decision making procedures as north Europeans do".*
>
> Thevenot (2008: 5-7)

Thevenot's description does seem to imply that a top-down, directive development process may have acted as something of a straightjacket to innovation. He notes, for example that

> *"a strong motivation is needed to convince the numerous stakeholders involved in any urban storm water management project of the effectiveness and sustainability of relevant 'Best Management Practices'".*
>
> Thevenot (2008: 5-7)

On the face of it, this appears as something of an admission of defeat when the aim was to develop an adaptive DSS. It seems debatable whether the involvement of such a wide variety of organizations in so many different countries and climatic settings actually works in favor of developing adaptive capacity - through consensus and standardization at each stage. The question might well be asked whether or not it might have been preferable to cultivate diversity of approaches, with periodic group reflection and sharing of experiences amongst different groups of innovators.

These observations should not be taken to imply that the DayWater Project has not made substantial progress. It has in fact created a DSS with numerous useful functionalities (library function, management function, analysis function and communication function). There is undoubtedly a web-based MCC tool available to its end users. However, it has many of the hallmarks of a smart "black box" when encountered by an uninitiated prospective user.

The critical question, in the context of this review, is whether the DayWater experience offers any useful guidance to the NIWA/Cawthron research team. In terms of technical, conceptual and analytical detail within the various DSS model components, the answer may well be yes, although that is not for this review to determine. But in terms of guidance on the concept development process, the answer is far from clear.

For the record, the DayWater MCC tool assists a user to select from a range of 15 structural BMP options to address storm water problems on a site with certain defined physical characteristics. There are 16 indicators across six categories (technical, environmental, Operational & Maintenance, Social & urban community benefits, Economic costs, legal & urban planning) to guide users' assessments, making the dimensionality of the DayWater tool considerably more complex than the tool envisaged in the NIWA/Cawthron research. Users choose how they weight the various criteria, with the sum of all weights equaling 100%. Performance scores for each criterion can either be across three bands (low, medium, high) as in the NIWA/Cawthron concept, or on a scale of 0-10. The interactive, computer-based tool allows users to engage in some individual experimentation or to generate several sets of results as the basis for discussion and group learning.

Beyeler (2008) provides an end-user perspective, representing one of the 14 core end users. Her paper makes it clear that she is not an uninformed member of the public, as indicated by her analysis of the needs for improvement in the management of urban stormwater in the face of unsustainable trends (p.166). Beyeler points out that the launch of the DayWater Project created high expectations, including (p.168):

- *"a tool easy to use by urban stakeholders (engineers, technicians, elected officials, planners, developers, etc.)*
- *a tool offering methodological advice, case studies, scientific, technical and financial information;*
- *a sharing of experiences, both good and bad;*
- *an incentive to think globally;*
- *a dynamic database;*
- *an adaptive approach allowing their request to be site specific."*

Beyeler then describes several experiences of interactions with the Hydropolis website which indicate how the DayWater tool has either been made more user friendly through end-user feedback, or might be further improved. It appears from her user perspective that the approach to developing the DSS has been technocratically driven by 'experts' in water management and information technology, but that the development process has not taken non-expert end users along with it and has therefore been experienced as somewhat inaccessible at times. At other times, she and other end users have been excited by their involvement. She argues for more direct involvement of end users in the testing and trialing of the DSS, but their involvement has to start at the beginning - articulating how end users experience problems and ask questions, and letting the DSS developers hear and respond to these.

Notwithstanding the apparent differences in research setting and scale between the NIWA/Cawthron research project and the DayWater Project, key lessons to be taken from the DayWater Project include:

- Managing expectations, particularly in deciding which user needs to address and how users express their needs,
- Working hard to establish a common language between scientists and non-scientists,
- Maintaining continual dialogue and feedback between DSS developers and users, and
- Making sure that the research team does not function in a fragmented manner.

Thevenot (2008:82)

3.2. Triple Bottom Line (TBL) Indicators

Overview

The 'triple-bottom-line' (TBL) term was first introduced by John Elkington in the late 1990s (see Elkington 1999) who described the approach as a way to measure and report performance against economic, social and environmental indicators. TBL indicators enable the understanding of scale and effects in each of the environmental, social and economic areas and have become an accepted international standard for robust full cost accounting of market and non-market natural resource values (Bebbington *et al.* 2007; Carpenter *et al.* 2009).

Since then the TBL approach has been more broadly defined as a way to capture the set of values and issues that ought to be addressed to avoid harm resulting from particular activities as well as to understand value under three pillars: economic, social and environmental value (Taylor 2005). TBL provides a framework for decision making in cases where sustainable development is the key objective. When applied in a participatory manner it forms a systematic and transparent decision support system for resolving multiple conflicting objectives, as described below (CSIRO Sustainable Ecosystems 2007).

TBL application for urban stormwater

The Multiple Objective Decision Support System (MODSS) developed by CSIRO uses MCA as an embedded method within a wider process to combine and display data which assists decision making (CSIRO Sustainable Ecosystems 2007).

Typical decisions that may involve a TBL assessment for urban storm water projects include:

- Assessment-related decisions for major storm water treatment development(s),
- Decision making relating to the location of new services,
- Examination of alternative strategies to deliver existing services,
- Asset management planning for storm water infrastructure, and
- Policy choices.

Figure 1 describes the MODSS process.

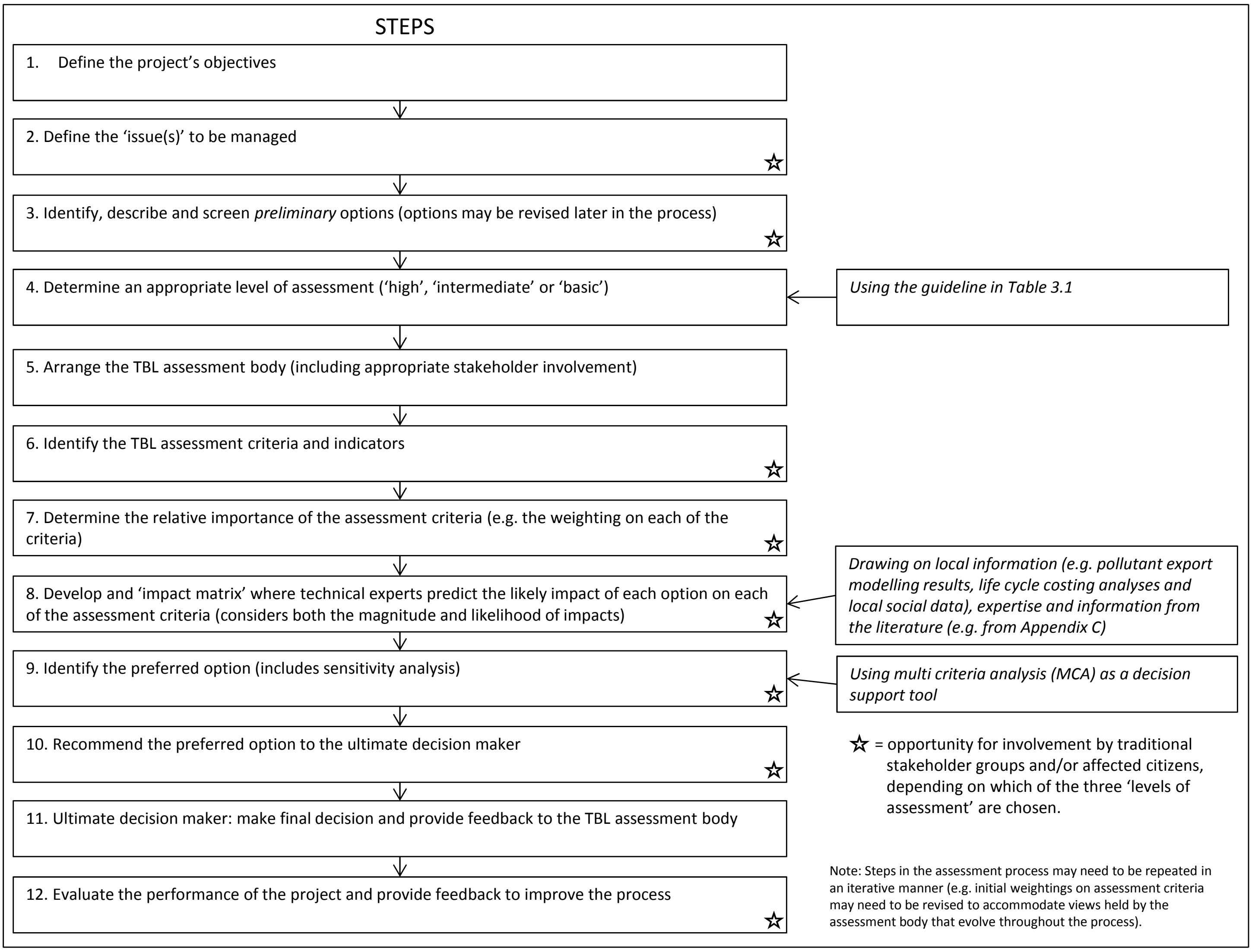


Figure 1. Process Overview Triple-Bottom-Line Assessment of Urban Stormwater Projects to Improve Waterway Health (Source: Taylor, 2005).

3.3. OECD Composite Indicator (CI)

This section of the report explains the mathematical basis of constructing composite sustainability indicators. The handbook by Nardo *et al.* (2005) describes a useful formalization of composite index construction, a method that can otherwise be prone to the criticism of being ad hoc. For researchers developing a composite index, it provides an excellent framework to follow, and would give credibility to projects whose development it guides. While it is founded on an example whose individual cases are countries, assessed on many aspects to build a composite index of their technological achievement, the methods described are applicable to a diversity of situations.

According to Nardo *et al.* (2005), the methods assume that the number of properties measurable for each case warrants the development of a composite index, and that the number of individual cases being assessed is many times larger than the number of properties. For example, if the cases being assessed are planning or policy scenarios, then for a composite index to be of value, those scenarios need to be evaluated in many dimensions (such as support by the community, cost of implementation, ease of policing, prevention of environmental impact, restoration of degraded sites), and there should be many different policy scenarios being assessed. It is also necessary to be able to make accurate and precise measurements of those scenarios. A composite index is usually based on quantitative rather than qualitative assessments of each case. The less quantitative these assessments, the less discriminatory the index will be.

Nardo *et al.* (2005) lays out a ten step process for the development of an index. The handbook is divided into several sections, the first outlining the ten steps involved in the development of a composite index, the second describing what makes a good quality index and how each step should be undertaken to maximize these aspects of quality. There follows a toolbox, describing options for the methods undertaken in each of the ten steps, their indications and contraindications.

The example presented in the handbook uses countries as the cases being evaluated, whereas in UPSW the cases would be urban planning policies. The handbook example uses assessments of aspects of industrial development as the indicators, whereas the UPSW project would define indicators around the effects of different policies. The handbook example describes the construction of an index called the Technology Achievement Index, whereas the UPSW project would develop an index describing the sustainability of urban planning policies. It is worth becoming comfortable with the distinction between the **cases** being evaluated, the **indicators** used to evaluate those and the **index** constructed from those **component** indicators.

This is shown in Figure 1 where three cases (scrolls 1,2,3 on the left hand side) are evaluated against three **indicators** (magnifying glasses A,B,C at the top) yielding nine scores (the plus/minus scorecards in the table). The three **component** scores for each **case** are gathered (stacks, right hand side), and combined in a construction process (the green shaded arrows) to an **index** (stars, right hand side). It is the process of combining individual component scores into an index that is described in the handbook.

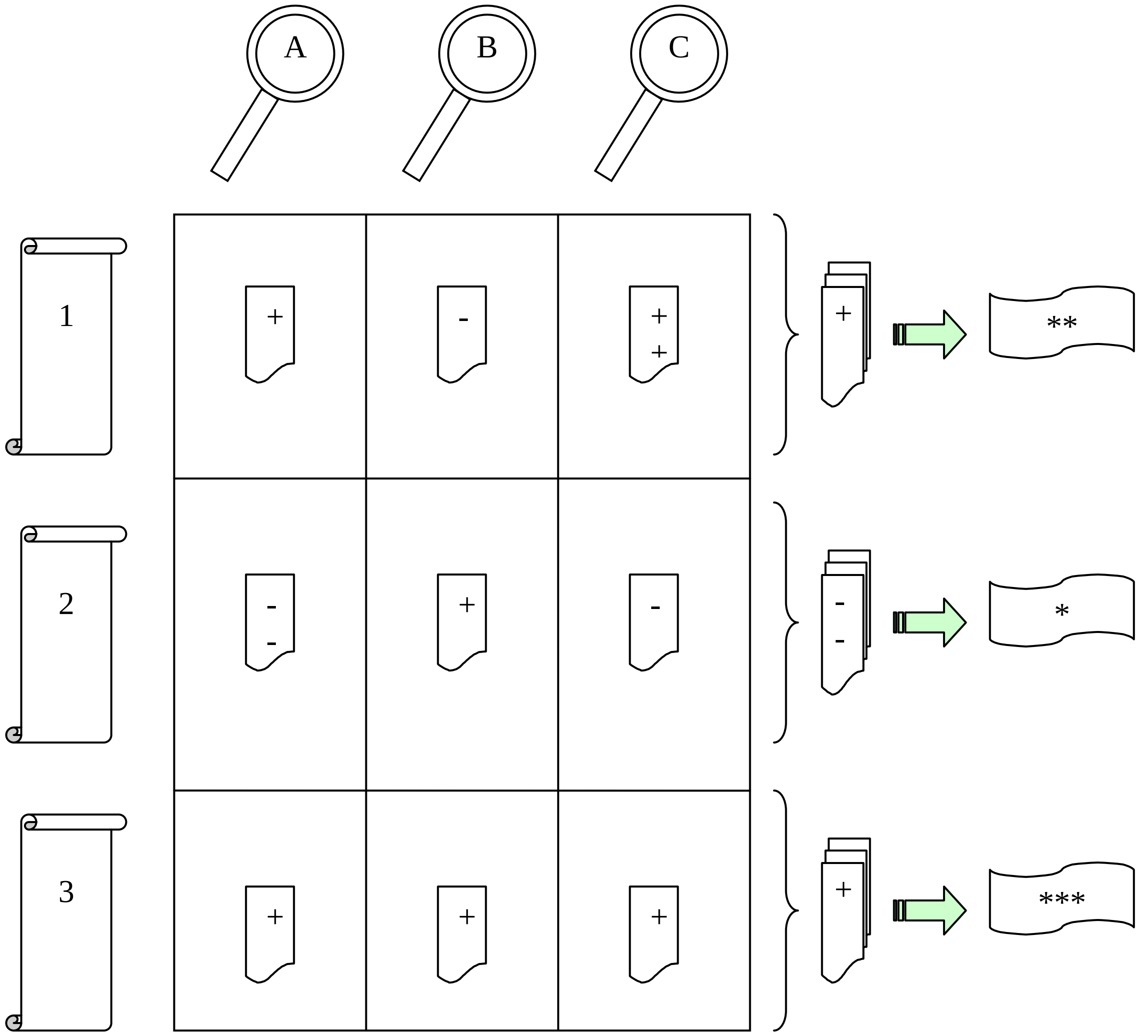


Figure 2. Composite Index Process: Cases 1, 2 and 3 assessed against Indicators A, B and C.

The framework described in the handbook is a useful formalization of a process which can otherwise be criticized (as described in Nardo *et al.*) as being ad hoc, or prone to arbitrary bias. The emphasis on documenting the process throughout is an important component of ensuring the acceptance of the composite index, and precluding criticism of a lack of transparency.

The handbook describes a range of methods suitable for different contexts or data types. If the developer was not familiar with the alternatives as presented in the handbook, they would use one by default, and not necessarily the most appropriate one. The developer may not appreciate the full implications, as they would not be working to an explicit framework. With no appreciation for the aspects of quality of an index they cannot realize what aspect they were aiming to improve with the method selected, and the method would then be less intentionally applied.

Despite the linear nature suggested by the ten step framework, it is worth becoming familiar with each section of the handbook before embarking on the process of developing a composite index. This is because it is important to consider what makes an index a high quality one, while implementing the methods for each step.

A search of google.scholar reveals that about 150 subsequent works cite Nardo *et al.* (2005), including further method development works as well as studies that have used their framework. These latter include studies on cardiac surgery, alcohol control policies, competitiveness of nations, and gender gap research, demonstrating that the approach has wider relevance than the comparison of nations presented. One study, the 2005 Environmental Performance Measurement Project (www.yale.edu/esi) does not cite Nardo *et al.*, yet its method framework echoes that of Nardo. (Their ESI calculation is made up of: variable standardization, variable transformation, multiple imputation, aggregation and weighting. There follow sections on quality, sensitivity analysis and statistical analysis.) This demonstrates that even when not following Nardo *et al.*'s framework explicitly, composite index developers often conform to the same methodological structure.

The first section of the handbook outlines ten steps to develop an index from component indicators. The ten step process is reviewed in more detail below. In its second section the handbook describes what makes a quality index. The definition moves beyond mere accuracy and precision, and incorporates more practical aspects such as timeliness and comprehensibility. These are to be considered throughout the development of the index, and Nardo *et al.* explain how each of the ten steps can contribute to, or impact on, the quality of the final output.

3.3.1. Developing a theoretical framework

The first step, according to Nardo *et al.* (2005), consists of purposefully planning the approach to the development of the composite index. The overall characteristic, or property, to be evaluated by the index is identified, and the component variables that may contribute to it are listed. These should be comprehensive, inclusive, and exhaustive, as some will be dropped out in step two to leave the most appropriate subset. They should aim to describe the many facets and elements of the property being quantified, and it does not matter at this stage if there is repetition or overlap. Documenting from the start the rationale behind the approach being taken will help clarify the method for the developer himself, and ensure it is well planned.

3.3.2. Selecting variables

The handbook describes the development of a composite index as including the components being combined (data, measurements) as well as the methods for combining them. The quality framework section describes what makes high quality data: there are a number of considerations when selecting variables to be the components of the composite. The rigorous application of high quality methodology to poor data will not yield a high-quality composite index. This section guides the selection of data components to ensure that the inputs to the process are not a limiting factor.

3.3.3. Multivariate analysis

This section of the handbook describes methods for investigating and rationalizing the component indicators to be combined into the composite index. It outlines a number of ‘traditional statistics’ analyses that can maximize the explanatory power of the index produced. That is, to ensure that the index can best differentiate between different cases, while not finding differences between similar cases. It will also help to ensure that the resulting index is as simple as it can be while still being effective (one aspect of the quality of an index), which is advantageous when explaining the index to an audience. However, the methods described (including principal component analysis and clustering such as by k-means) have the potential to obscure the component scores, which is at odds with the aims of clarity and decomposability stated elsewhere.

3.3.4. Imputation of missing data

When dealing with real collected data, missing values are an inevitable reality. If one of the cases being evaluated has no score against one of the criteria, this will bias its index score. This section of the handbook describes how missing data can be tolerated, by filling any gaps using whichever method is most appropriate to the nature of the data. The implications of the various options available are outlined with regard to the different aspects of the quality of composite indices.

3.3.5. Normalisation of data

Individual measurements or asessments (countries in the handbook example) may be measured against a number of different criteria, each of which may have different scales of measurement. For instance, area in km^2 may have values in the range of hundreds of thousands, whereas years since suffrage may have values in the range of tens to hundreds. In order for measurements on such differing scales to be combined without the larger numbers swamping the smaller numbers, they must be brought into a common range. This is done by normalising, and again there are different techniques, which are discussed along with their implications in the handbook.

3.3.6. Weighting and aggregation

According to Nardo *et al.*, the issue of weighting “must be considered”, as in the absence of any more intentional weighting, a flat weighting is in effect being selected by default. While equal weighting is not necessarily a concern *per se*, the weighting mechanism offers an opportunity to impart more significance to some variables than others. For instance it may be decided that while data on both land area and years since suffrage are available for all countries, for the index being developed the land area is less important than the number of years since suffrage. This relative importance can be implemented via a differential weighting of the component indicators.

3.3.7. Robustness and sensitivity

The various processes in the index methodology should each have an effect on the index created, and partially influence how the many cases compare one to the other. While the techniques described in the handbook's discussion of robustness and sensitivity do not directly affect the index value, or the comparison of cases, they do allow the influences of the other steps to be analyzed and described. This is important for the reporting of the construction of the index. For instance, the weighting selected for the components of the composite will affect each case's final score. While the weighting regime should be transparent and reported, it is also useful to demonstrate the sensitivity of the final scores to different weighting regimes.

3.3.8. Links to other variables

Again, while the method described under this section does not directly affect the index or case comparison directly, it represents a valuable evaluation of the index and its construction, by comparing it to existing measures of related properties of the same cases. If a well-accepted comparison of the cases exists (perhaps against different criteria), then showing a correlation between the accepted score and the newly developed index lends credibility to the new index.

3.3.9. Back to the details

The ninth step seems somewhat at odds with the philosophy of simplicity provided by a composite index. Nardo *et al.* recommend exposing the component measurements that make up the composite, to provide more detail on how each case gained the index score that it did. This allows the constructor of the index to verify that component scores are real, allowing the reader to see the basis for a specific case's score on the index, and allows stakeholders or those responsible for an individual case to identify their strengths and weaknesses as measured by the index. That is, if the cases being investigated were schoolchildren, and the composite index was built upon their grades in each subject, the composite index would give them an idea of their overall performance, while the breakdown would reveal the subjects in which they could improve for most benefit to their overall score.

3.3.10. Presentation and dissemination

Part of the justification for presenting a composite index rather than all individual component scores is to improve the accessibility, or the ability of the contained information to be understood by a target audience in a short time. The presentation of the index and comparison of various cases is key to this understanding, and the handbook provides a number of options. It is however rather light in this section, and there are many more display options than it details. However the principles that it recommends for consideration in presenting results are valuable and will apply to graphics other than those considered by the authors.

3.3.11. Summary

The handbook presents a useful distillation of the process involved in the development of a composite index. It clarifies complex mathematics that would not otherwise be transparent to the naïve developer. However, Nardo *et al.* (2005) make no comment as to determination of the appropriateness of composite indices for specific studiers. Some of the processes in the steps require a large number of cases to ensure the reliability of their outputs, and the implications of skipping this step if few cases are available may warrant investigation before the adoption of a composite index methodology.

3.4. Multi-criteria analysis (MCA)

A number of variants of the MCA approach have been developed over time. In this section of the report these methods are listed and reference to more detailed explanations are provided. Additionally, several software tools were identified that may have some benefits in realization of the UPSW aims and a suggested way of communicating scenarios and outcomes with participants in the process is described. After a short summary of definitions, an outline of the process steps involved in MCA (see Section 3.4.2) is presented. A diagrammatic view of the MCA framework is provided that further explains the connections between the steps outlined in Section 3.4.2 MCA Process.

3.4.1. Definitions

MCA is also referred to by several other names as noted by Hajkowicz and Collins (2007) and Proctor & Qureshi (2005). They include:

- MODS (Multiple Objective Decision Support)
- MADM (Multi-Attribute Decision Making)
- MCDA (Multi-Criteria Decision Analysis)
- MCE (Multi-Criteria Evaluation)
- DMCE (Deliberative Multi-Criteria Evaluation)

Integrated assessment is "the combination, interpretation and communication of knowledge from diverse scientific disciplines from national and social sciences to investigate and understand causal relationships within and between complicated systems" (IPCC, 2001 in Holman *et al.*, 2008 Regional Integrated Assessment Software).

Aims of integrated assessment

- Disciplinary integration and provision of new information about complex systems that enables or improves decision-making, and
- Provision of better understanding of socio-economic scenarios, is able to provide sensitivity analysis particularly re multiple interactions, feedbacks and potential adaptive responses (Holman *et al.* 2008).

3.4.2. MCA Process Steps

The overall aim of MCA is to choose between a finite number of options by using a set of criteria to judge each option. Methods are applied in order to rank options on how well each option satisfies the criteria. The MCA process stages are briefly outlined as per Hajkowicz & Collins (2007):

1) Choose decision options: explore all possible options and then narrow down to a finite (discrete) choice set.
2) Select evaluation criteria: ensure that all criteria are relevant and remove any highly correlated criteria.
3) Obtain performance measures: experts or other environmental/economic models usually including both quantitative and qualitative measures.
4) Transform to commensurate units: for meaningful inclusion in the utility function.
5) Apply criteria weightings: to adjust the importance of each criteria as these are rarely equal.
6) Rank options: combine criteria weights and performance measures to rank the options being considered.
7) Conduct sensitivity analysis: improve model strength and use ranking algorithms to ensure input assumptions provide robust results.
8) Make a decision: use some level of human judgment, particularly for issues that cannot be adequately catered for in the MCA model.

There are various MCA techniques available and studies show that these methods produce similar results. Hajkowicz & Collins (2007) review of MCA for 113 published water resource planning and management MCA studies from 34 countries found agreement between techniques, which they categorized as:

- Fuzzy set analysis
- Pair wise comparisons
- Outranking methods
- Distance to ideal point

- Multi-criteria value function
- Weighted summation/multiplication.

More detailed explanation of various MCA methods is available in Figueira *et al*. 2005.

Figure 3 provides a visual representation of the MCA options and flow of information and decision steps.

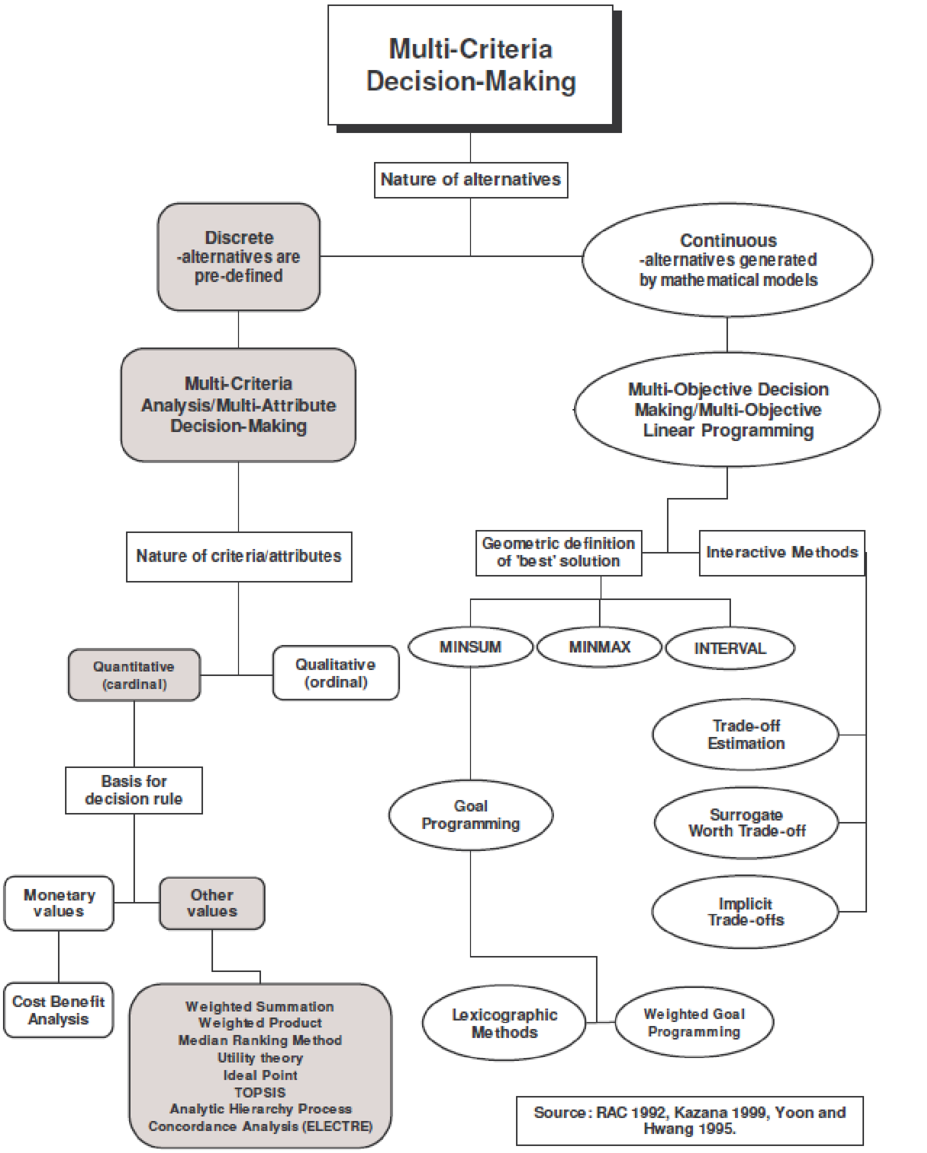


Figure 3. Overview of MCA process and related taxonomy (Source: Proctor & Qureshi 2005, original sources noted in diagram above).

3.4.3. *Software*

There is a variety of software programs available and easily accessible for MCA *e.g.* **DecisionLab** (Proctor & Dreschler 2003). **Regional Integrated Assessment** software tools may be useful and perhaps warrant further investigation (Holman *et al.* 2008). Additionally, spatial decision making for MCA may benefit from application of the CORMAS available at http://www.cormas.cirad.fr (Perez 2010). **MCAT** (Multi Criteria Analysis Tool) has been specifically designed for water management; the software and user guide can be accessed at http://toolkit.ewater.com.au/Tools/MCAT/documentation (CSIRO Sustainable Ecosystems, 2007).

3.5. Strengths and weaknesses: multi-criteria processes versus composite indices

Section 2.2.1 describes strengths and weaknesses of MCA processes from expert and stakeholder/participant perspectives. This section builds on that material to generate a comparative assessment of the strengths and weaknesses of the Composite Indicator (CI) and Multi-criteria Analysis (MCA) from a variety of factors derived from commentary by Baines and Morgan (2010). Table 1 presents the outcome of the comparative analysis.

Table 1. Strengths and Weaknesses: Multi-criteria processes versus Composite Indices.

Contrast Factor	CI	MCA
Frequency of use in water management	(W) none detected, however, there may be case studies underway that are yet to be published. (W) recently defined process, first publication 2002 (Nardo *et al*. 2002).	(S) 113 published cases in 34 countries (see Hajkowicz & Collins 2007). (S) Method has been applied since 1973 and has been significantly improved over time.
Bias	(S) avoids stakeholder bias by excluding stakeholders from the weighting and ranking process (Saisana & Tarantola 2002).	(W) Focus groups may have self-selection bias leading to confirmation bias (Dobes & Bennett 2009).
Complexity	(W) Two extra process steps than MCA.	(S) Less complex than CBA (Dobes & Bennett 2009).
Stakeholder participation	(W) Stakeholders only in the indicator selection process.	(S) Stakeholders included in all steps of the process. Note: this is especially true with Deliberative MCA where objectives, rankings and weightings are stakeholder derived. (W) Process can be time intensive and result in stakeholder burnout (Proctor & Qureshi 2005).
Stakeholder equity	(S) Facilitates communication with general public (*i.e.* citizens, media, *etc*.) and promote accountability with stakeholders.	(W) In some case, process is centralised, technical and highly political nature. (W) Barriers to some stakeholder participating may be due to lack of facilitator experience and differing communication styles (Proctor & Qureshi 2005).
Verification (correct model build)	(W) By developers only.	(S) By developers and stakeholders.
Validation (building the "correct model", an acceptable level of accuracy in model predictions)	(W) By developers only.	(S) By developers and stakeholders.

Contrast Factor	CI	MCA
Sensitivity analysis (extent of variation in predicted performance)	(W) Poorly constructed / explained indicators open to misinterpretation. (W) Stakeholders not privy to sensitivity analysis.	(S) Decision making can be explored at greater depths providing greater insights into the nature of the problem, unravel complexities and can provide recommendations for future analyses (Proctor & Qureshi 2005).
Decision support	(S) Composite indicators summarise complex or multi-dimensional issues for supporting decision-makers. (S) Facilitates rankings on complex issues where the aim is benchmarking. (W) May result in misleading, simplistic or inappropriate policy conclusions particularly if difficult to measure performance dimensions are ignored (Saisana & Tarantola 2002).	(S) Yes as ranks/scores performance of alternative decision options against multiple criteria which are typically measured in different units?
Base case	(W) no.	(S) yes, in many cases.
Dimensionality	(S) Multiple dimensionality – use of weights for comparisons.	(S) Multiple dimensionality – use of weights for comparisons.
Multi-stage	(S) 10 steps. (W) many steps are controlled by developers only.	(S) Yes – first stage often expert panel, later stages can incorporate stakeholders (Proctor & Drechsler 2003).
Reproducibility	(S) Yes.	(W) Some limits since each case individually designed (Dobes & Bennett 2009).
Quantitative & Qualitative	(W) Qualitative data are not included (Saisana & Tarantola 2002).	(S) Able to use both quantitative and qualitative data (Hajkowicz & Collins 2007).
Presentation/dissemination	(S) Specific process step to ensure communication of results, see Figure 2 below.	(W) Not specified in MCA process. (S) Stakeholders involvement in dissemination.
Interpretation	(S) Composite indicators easier to interpret than trying to find a trend in many separate indicators. (W) Combination of various indicators into CI may disguise serious failings in some dimensions which may increase the difficulty of identifying options for remedial action.	(S) Human judgment required to some extent as the outcome is not a single answer (Hajkowicz & Collins 2007).
Origin	Statistical and social sciences.	Social and ecological sciences.

3.6. Complementary approaches

During the literature review we also looked to identify alternative indicator development methods. Four alternative or complementary approaches to CI and MCA were identified and are outlined below.

3.6.1. Mediated modeling

Mediated modeling is an inclusive approach to model development that can reduce the risk of criticism by stakeholders by involving them in model generation (*e.g.* van den Belt 2004; van den Belt *et al.* 2006). Having participated in the development of the model (both structure and parameterization), potential critics have ownership and understanding of conclusions drawn from its use.

In this approach the "modeling" is of an illustrative, explanatory nature rather than a quantitatively predictive one. That is, the model developed is a representation of the system being studied: the stocks or resources, their interactions and the modulation or regulation of those interactions. This representation can help people to understand phenomena that the real-world system exhibits, can reveal the key drivers of those phenomena, how they might be exacerbated or moderated, the feedbacks and consequences of actions. The models are not intended to be of the accuracy required to make strict quantitative predictions.

The "mediated" aspect describes the requirement for a modeling leader to guide and arbitrate the process of stakeholder group contributions to the modeling process. These stakeholder contributions consist of their understandings of the system being studied and insights or observations they may bring that can guide the structure or the parameterization of the model. Mediation is required to ensure all participating stakeholders have confidence in the contributions that are incorporated into the model. Consensus understanding is reached through dialogue led by the mediator before any contribution is incorporated. Where consensus cannot be reached due to the diversity of stakeholders' perspectives or world views, the effect of alternative understandings on the overall system behaviour can be investigated by substituting alternative model components, each representing one of the various understandings.

Mediated modeling does not prescribe a simplistic approach to modeling. The "modeling" implies use of mathematical/logical methods (as opposed to narrative, spiritual, or metaphysical epistemologies). A participatory approach could be applied to developing a collaborative narrative understanding, but mediated modeling implies the development of a model which may be: a physical model, a statistical model, a system dynamics model, or Bayesian Belief Network model, or some combination of them. In any of these scenarios, the representation of stakeholders' knowledge and concerns facilitates rapid model development where a lack of published or validated data might limit the potential for a single authority to develop a literature-referenced model. However, for some participants the requirement to incorporate their (*e.g.* spiritual or anecdotal) knowledge in a logical framework could represent a culturally inappropriate imposition.

The mediator also has responsibility to ensure that contributions to the model are attributed, to record that consensus was reached on their incorporation and to detail where any exceptions to that consensus could not be resolved. This constitutes a claim that all stakeholders were satisfied with the representation of the study system, and that they have ownership over results generated, and reduces the likelihood for those results to be subsequently challenged by the stakeholder groups involved.

It is therefore critical that the development process involves as many of the stakeholder groups as practicable, and that they are informed from the start of how the model will be used. It is important that the limitations of the modeling approach are acknowledged during use and presentation of the model and any results derived from it.

3.6.2. Drivers–Pressure–State–Impacts–Response (DPSIR)

The DPSIR (Drivers–Pressure–State–Impacts–Response) framework is used for Australian State of the Environment reporting by Commonwealth and State governments (see http://www.environment.gov.au/soe/index.html). This is generally a framework for reporting the status of environmental indicators. It should be noted that this type of framework does not necessarily capture all desired indicators. Authors of State of Environment reports have commented on the lack of a nationwide consistent, comprehensive indicator system (Ward and Butler 2006).

3.6.3. Bayesian Belief Networks (BBN)

The OECD approach to developing composite indicators (Nardo *et al.* 2005, described in section 3.3.1) requires the development of a conceptual model of the system under consideration. On the basis of that model, further models may be developed to generate and combine data from variables that map the system to form a composite indicator.

A Bayesian Belief Network (BBN) is a specific implementation of the wider family of Knowledge Network models. Generic network models provide a graphic conceptual map of the linkages between stressors and outcomes in a system. BBNs add a probabilistic dimension that provides an avenue to take account of uncertainty and to generate information on the levels of influential variables in the system. BBNs use mathematical and statistical concepts to generate forecasts of the probabilistic strength of linkages in a system and of the likelihood of outcomes of interest to decision makers. BBNs may be developed and operated in a variety of contexts. They are able to be informed by expert panels, data sourced in the literature, model sourced data and knowledge held informally by stakeholders and decision makers. As such they have application in mediated modeling processes in the context of ecosystem management and adaptive management processes. It follows that they may also have a key role in the development of expert software/decision support systems (Ticehurst *et al*. 2007).

BBNs address the governance and management requirements described in Section 2.1. Because of their flexibility they are able to be used in consultative processes such as mediated modeling, with the BBN contributing visual, conceptual, and analytical strengths while accommodating uncertainty in the process. Flexibility in specification of the final decision nodes allows incorporation of component indicators in either interval or ratio forms: the latter has particular relevance to the “strong” sustainability argument for identifying targets, limits and thresholds against which indicator levels are assessed.

Castletti and Sonini-Sessa (2007) provide a technical summary of BBN, noting the ready availability of ready-to-use software with which BBN models can be designed and implemented.

There have been a number of applications of BBNs to estuarine and freshwater management in New Zealand. Using Netica software, Quinn *et al.* (2009) modeled catchment processes at Bog Burn in Southland. Quinn *et al.* (2011) developed a BBN to analyze water supply options in the Canterbury Water Management Strategy for the Culverden Basin/Hurunui system. Hume *et al.* (2009) developed a BBN to inform sediment management in the Mahurangi Harbour and associated catchments. These implementations used Netica 4.02 developed by the Norsys Software Corporation of Vancouver, Canada.

Figure 4 provides an example of the effectiveness of a BBN in capturing the complexity of freshwater system.

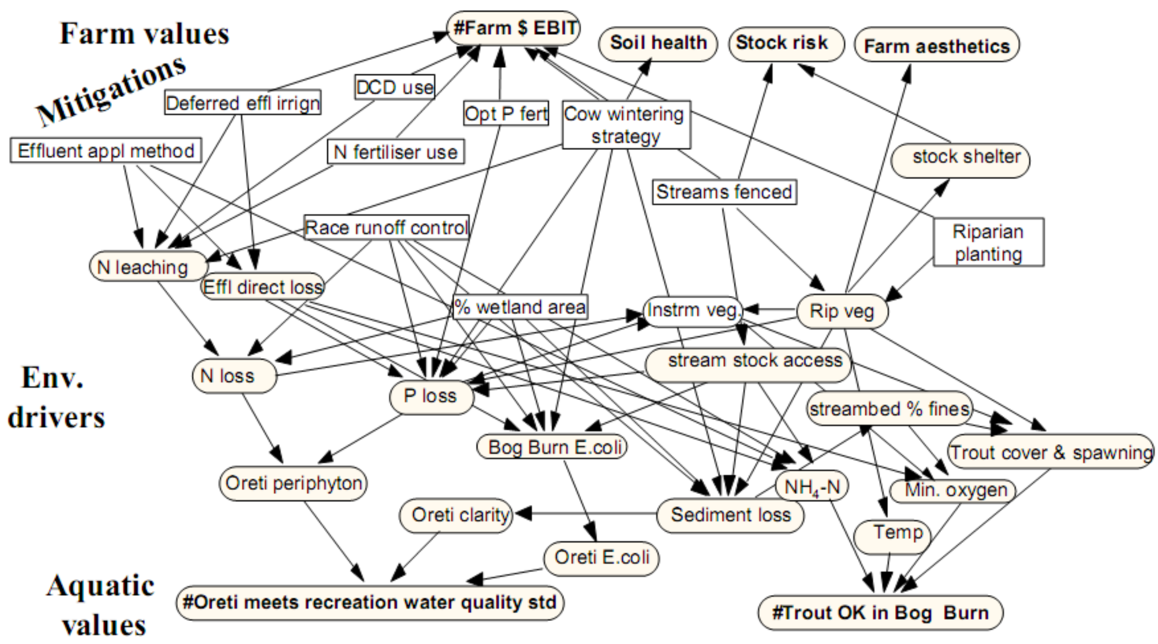


Figure 4. BBN conceptual map of linkages in the Bog Burn system. (Source: Quinn *et al.* 2011).

3.6.4. *Other approaches*

A number of other techniques have been employed to generate conceptual maps of environmental management and policy settings. They include system dynamics (Maani & Cavana 2000), Meta models, and coupled component models (Leitcher & Wiederman 2004). However, these techniques suffer from complexity and inability to effectively deal with uncertainty, and fail to provide managers with a forum to explore human/nature interactions holistically (Ticehurst *et al.* 2006). In contrast BBNs are able to model uncertainty to promote stakeholder participation, and to integrate information from social, cultural, economic and environmental domains.

4. THE UPSW COMPOSITE INDICATOR SYSTEM

This section of the report describes a system for sustainability indicators for the UPSW DSS. This indicator system combines elements from each of the processes reviewed in Section 3. The technical aspects of the recommended system follow that outlined by Nardo *et al.* (2005), which we characterize as the OECD or CI approach. This forms the core of the method.

Elements are selected from the CI approach based in recommendations from the governance context, the key features of existing systems, and multi-criteria analysis, to form the design of the sustainability index system.

The subsequent sub-sections examine in turn the features of the CI method:

- The conceptual map,
- Identification of key elements,
- Normalisation of raw scores,
- Incorporating limits and targets,
- Derivation of weights,
- Aggregation schemes,
- Reporting, and
- Methods for sensitivity analysis and validation.

The research framework of UPSW is a software development context. It follows that techno-centric approaches will form the core of any method. It is recommended that a variant of the OECD method (Nardo *et al.* 2005) is implemented because that source is respected as an effective technical summary and handbook for composite indicator developers that:

- Prescribes a sequence of development steps and tests to derive an effective index that produces meaningful, consistent results for decision makers,
- Elaborates options within each development step.

The approach follows closely the overview described in Figure 4. A Bayesian Belief Network generates the conceptual map, specifying linkages between stressors and outcomes and generating the data that forms the raw scores on variables that are the precursors to the sub-indicators. Those sub-indicators are then combined with weights to form the composite index. To facilitate combination, the raw scores are transformed through a normalisation process that expresses the sub-indicators on a common scale spanning the interval [0 -100]. Two aggregation schemes are offered: arithmetic mean and geometric mean combination. The key difference lies in the degree to which favourable levels of one sub-indicator can compensate for less favourable levels of another sub-indicator in the final composite indicator score. The arithmetic mean is the aggregation device in a linear additive combination model with full compensation between variables. The geometric mean creates a non-linear multiplicative combination model that limits compensation between variables. Key to discrimination between the two aggregation schemes is the capacity to consistently rank development options.

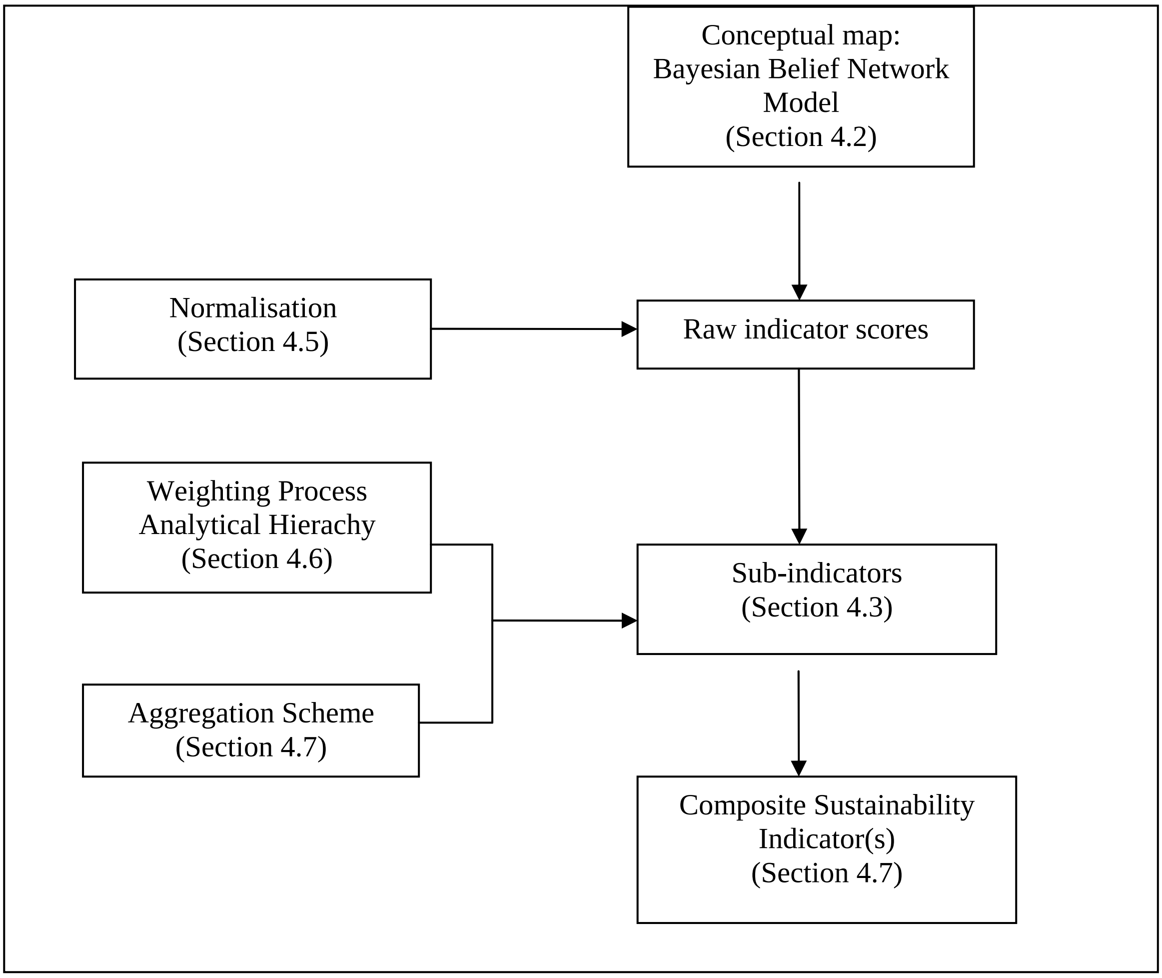


Figure 5. Overview of recommended sustainability index system for UPSW.

4.1. Key contributions from the governance/management context

Key contributions from the OECD approach that arise from the governance context (Section 2.1) are:

- Identification of relevant perspectives on sustainability: “weak” versus “strong”,
- Identification objectives for assessment and decision making,
- Clarification of trade-off rules,
- The role of the DSS tool in adaptive processes, and
- Identification of the relationship between those involved in assessing and those involved in decision making.

In particular,

- Stakeholder involvement is crucial in derivation of a conceptual framework and selection of relevant criteria for sustainability assessment,
- A clear definition and incorporation of the objectives and goals of the decision makers using the DSS into its design, in particular the preferred sustainability definition is required,
- A clear understanding of decision makers and stakeholders preferences in respect of trade-off rules is required, and,
- Thresholds, limits and targets should being specified and feature in indicator system design.

Definition of objectives and goals in turn leads to specification of targets and limits for the variables that are the raw data for the composite indicator process. An interactive weight specification process meets the requirement to incorporate stakeholder and decision maker preferences for trade-off rules.

In order to achieve these goals, the modified OECD CI process will:

- Employ a Bayesian Belief Network as a conceptual mapping technique that allows stakeholder participation. When used in this way, the BBN process is itself a form of mediated modeling,
- Create component indicators from raw data scores that incorporate limits, thresholds and targets where relevant,
- Adopt appropriate methods for normalisation of component indicators given the scale on which they are measured, selected from either ratio or interval scales,
- Use an interactive technique to allow stakeholders and decision makers to participate in the derivation weights for aggregation of the individual component indicators to a composite indicator, and,
- Employ an aggregation method that captures preferences for trade-offs *e.g.* compensability and that shows the highest level of consistency in ranking development options available (Section 4.7.4).

4.2. Conceptual Map: Bayesian Network Model

In contrast to alternative methods described in Section 3.6, Bayesian Belief Models (BBN) are able to model uncertainty, to promote stakeholder participation, and to integrate information from social, cultural, economic and environmental domains. Section 3.6.3 has described the suitability of BBNs for the task of constructing conceptual maps of catchment and estuarine and freshwater processes and generating data that forms the basis of composite indicator processes. The process for defining the BBN may include expert and stakeholder inputs so that the pathway between catchment storm water stressors and the CI is participant driven.

4.3. Sub-indicator selection

Sub-indicators are the immediate precursors of the composite index. Sub-indicators are selected from a mix of theory and expert, stakeholder and decision maker sources. The Bayesian Belief Network process described in Section 4.2 ultimately defines the final nodes of the network which form the data sources for the sub-indicators.

4.3.1. Biophysical sub-indicators

The UPSW project aims to evaluate urban development effects on both stream and estuarine receiving environments. Two sets of (probably different) biophysical sub-indicators are required to derive ecological health indicators for these two types of waterbody.

Sub-indicators of stream ecological health could include:

- Physical habitat characteristics (in-stream and riparian),
- Water quality, and
- Characteristics of community and/or species of aquatic biota.

Sub-indicators of estuarine ecological health could include:

- Sedimentation rates,
- Sediment quality, and
- Benthic health characteristics.

4.3.2. Socio-economic sub-indicators

Preliminary stakeholder engagement in recent choice modeling studies are useful as a guide to identifying final nodes in the BBN framework that contribute data to socio-economic scores. Kerr and Sharp (2004) identified the following attributes in their choice model project around community preferences for Auckland streams:

- Channel shape,
- Water clarity,
- Abundance of native fish species,
- Fish habitat (reflecting underfoot conditions in stream beds), and
- Riparian vegetation.

Batstone and Sinner (2009) found the following attributes to be the key factors underpinning community preferences for coastal environments:

- Underfoot conditions,
- Water clarity, and

- Ecological health (reflecting fish abundance).

4.3.3. Resilience indicators

This section of the report provides a brief overview of the Resilience concept. The resilience literature is rapidly expanding. Further detail may be obtained from Walker and Salt (2006) and the various papers published by Hollings, Carpenter, and Folke on the subject.

An emerging approach to understanding sustainability is Resilience. Resilience is the capacity of a system,

> *" to absorb disturbances and to re-organize while undergoing change so as to still retain essentially the same function, structure, identity, and feedbacks"*
>
> Folke (2006:259)

This definition emphasizes maintenance of a system's adaptive capacity – the capacity to adapt to the effects of stressors while maintaining a desirable state, or, by implication, the capacity to change, effecting a transformation to more desirable states.

Applied to ecosystem management, the resilience concept relies on system specification as a combined socio-ecological system. The feedback processes between the two are consequential: it is not possible to manage social and ecological systems in isolation. This way of thinking facilitates understanding of how societies, economies and ecosystems may be managed in combination to promote resilience - maintaining the capacity of the system to absorb disturbance without changing state. This approach to sustainability goes beyond weak sustainability (Section 2.1.1) to the strong sustainability definition in which desired goal states are specified. Resilience thinking allows targets for the management of key influences to be specified in the context of the dynamics of system under consideration.

The following synthesis of resilience thinking summarizes the principles necessary to understand resilience in estuarine and freshwater systems:

- Estuarine and freshwater systems are viewed as social-ecological systems that integrate ecosystems with human society.
- Systems theory posits that systems have the capacity to exist within one or more equilibria or stable states, not all of which are favorable from an anthropogenic perspective.
- Alternate stable states potentially may exist within estuarine and freshwater systems.
- Estuarine and freshwater systems' properties can vary significantly within any one stable state.
- Estuarine and freshwater systems can exhibit considerable spatial and temporal variability at different scales within a stable state.

- Thresholds exist in estuarine and freshwater systems which may be tipping points between alternate stable states.
- The term regime shift is used to describe the shift in a system, reflected in key parameters, between alternate stable states.
- Thresholds may exist at different scales. However not all thresholds automatically imply a regime shift in the system.
- Slow variables such as land use change are influential in driving regime shift.
- Estuarine and freshwater systems may cycle through recognized adaptive loops: their location in this process determines their form and function.
- Managing estuarine and freshwater social-ecological systems requires the capacity to adapt to and to influence change in order to achieve desired system states that are the goals of intervention policy.

Parsons *et al.* (2009)

The adaptive cycle has been described in terms of four phases identified through the study of diverse ecosystems worldwide (Gunderson and Holling 2002):

- Rapid growth,
- Conservation,
- Release, and
- Reorganisation.

The simplest transition pathway is a cycle through stages defined by each of these terms, although alternate pathways have been observed. Transition pathways are possible, and have been identified between all phases except from release and reorganisation phases directly to the conservation phase. Figure 6 demonstrates the concept, with the smaller internal arrows representing possible alternative pathways.

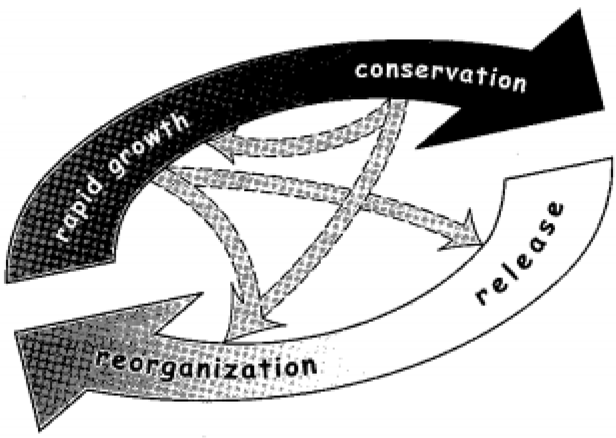


Figure 6. The Adaptive Cycle (Walker and Salt, 2006).

Identification of the resilience state of a system, its capacity to retain current function and condition, may be achieved by creating a two dimensional plot which relates levels of so-called "fast" and "slow" variables. Carpenter *et al.* (2001) describe the measurement of the extent of resilience in a lake system by identifying equilibria in terms of sediment phosphorous (the slow changing variable) generated by land use and its corresponding fast changing variable, water phosphorous. (Figure 7.)

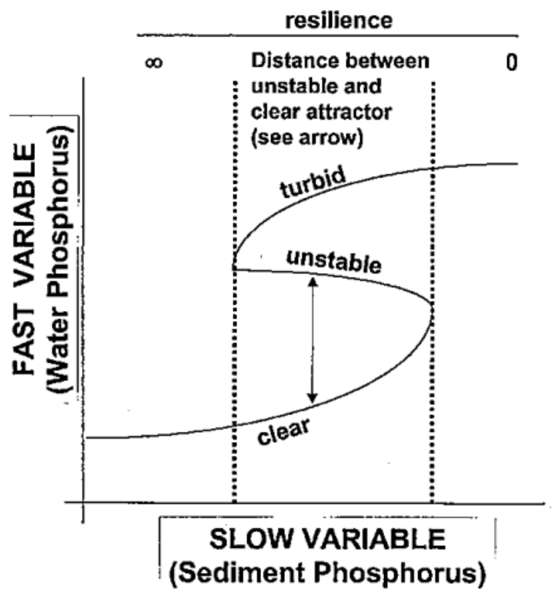


Figure 7. Resilience measurement concept in lake ecosystem (Carpenter *et al.* 2001).

Two potential stable states, or attractors, are shown: the desirable clear water state, and the undesirable turbid or eutrophic state. Specification of the exact dimensions of the attractor system from field data is non-trivial. However soil and sediment phosphorous levels are readily measured. The key implication from Figure 7 is that resilience of the clear water, oligotrophic state of the system is inversely related to levels of sediment and soil phosphorus. Phosphorous loads may therefore be one important indicator of the resilience of the lake system. It follows that measures derived in other contaminant and sediment loads that contribute to the eutrophication process in the lake may be useful as linked indicators of increasing or decreasing resilience.

Carpenter *et al.* (2001) list the following system indicator variables for a lake social-ecological system consisting of agricultural activity (stock grazing) in a lake catchment:

- Net social utility of the system,
- Intensity of agriculture,
- Regulation of agriculture,
- Number of attractors,
- Clear water attractor size (reflected in soil phosphorous),
- Vulnerability to shocks,
- Water quality,

- Sediment phosphorous, and
- Soil phosphorous.

This example forms a template for development of resilience indicators for the estuarine and freshwater systems that the UPSW project aims to evaluate. The resilience concept provides a framework with a reasonable pedigree to move from system definition focused on the biophysical receiving water bodies to a whole system focus which spans the entire social-ecological system of human urban development and associated catchment and downstream effects.

4.4. Data reference points

Let RS be the raw score on a final variable or node in the Bn, and i be the number of final nodes. Then RS_i is the raw score on node i.

RS_i has a number of states or levels that occur in its range of possible levels. Let Min_i be the minimum possible level for RS_i, and Max_i be the maximum possible level for RS_i.

There are two further reference points that may occur in the range of RS_i. These reference points are established by stakeholders and by decision makers. They are a specification of the permissible limit (L_i) a score may take, or a target score (T_i) set by policy or regulation. Some possible relationships between RS_i, Min_i, Max_i, L_i, and T_i are shown in Figure 8. While the reference points are fixed, RS_i may lie anywhere in the interval between Min_i and Max_i and can change over time.

$$Min_i \quad < \quad RS_i \quad < \quad T_i \quad < \quad L_i \quad < \quad Max_i$$

Figure 8. Key reference points for raw indicator scores.

4.5. Normalisation of data

The normalisation process converts RS_i to sub-indicator S_i. The normalisation scheme described in this report is referred to as the re-scaling or "min-max" approach in the literature (Nardo *et al*. 2005). Use of this scheme is a necessary precursor to the use of linear additive and geometric models for aggregation of the resulting sub-indicators to form a composite indicator (Section 4.7).

The goal of the normalisation process is to create a common interval scale across all sub-indicators (S_i) that contribute to the composite indicator (CI).

The normalisation proceeds in two phases:

1) Creation of a ratio that reflects the position of RS_i in its range.
2) Conversion of that ratio to a position on the interval [0 – 100], by multiplying the ratio by 100.

The denominator for the normalisation ratio depends upon how the reference points have been defined. Three possible cases emerge from the specification of the range and reference points for $RS_{i.}$:

1) No limits or targets are specified.
2) A limit has been specified.
3) A target has been specified.

Case 1: No limits or targets

Where there are no limits or targets, the sub-indicator SI^{NLT} is defined by its location in respect of the possible minimum and maximum values.

$$SI^{NLT}_{i\,i} = \left(\frac{RS_i - Min_i}{Max_i - Min_i} \right) * 100 \qquad \text{Equation (1)}$$

Where SI^{NLT} is the normalised sub-indicator in the case of no limits or targets, denoted by the superscript “NLT”, derived from the raw score RS_i .

Case 2: A limit has been specified

Where a limit has been specified, $SI^{L}_{i\ i}$ is defined by its location in respect of the possible minimum and limit values.

$$SI^{L}_{i\ i} = \left(\frac{L_i - RS_i}{L_i - Min_i} \right) * 100 \qquad \text{Equation (2)}$$

Where $SI^{L}_{i\ i}$ is the normalised sub-indicator in the case of limits specified, denoted by the superscript “L” derived from the raw score $RS_{i.}$

Case 3: A target has been specified

Where a limit has been specified, SI^{T}_{i} is defined by its location in respect of the possible minimum and target values.

$$SI^{T}_{i\ i} = \left(\frac{T_i - RS_i}{T_i - Min_i} \right) * 100 \qquad \text{Equation (3)}$$

Where SI_i^T is the normalised sub-indicator in the case of a target being specified, denoted by the superscript "T".

4.6. Assigning relative weights to indicator components

Indicator weights are an important component of the aggregation process that draws together information from a number of variables (SI_i) to deliver a composite indicator (CI). Depending on the techniques by which they are derived, weights have the potential to allow both physical and psychological factors to be incorporated in indicator processes (Saaty 1987; Nardo *et al.* 2005).

Statistical techniques such as principal components analysis (PCA) create weights that correct for overlapping information in two or more correlated factors, as opposed to reflecting the theoretical or assessed importance of individual SI_i in a bundle of SI_i that contribute to a CI.

The following techniques are available to create indicator weights:

- Principal components analysis,
- Linear programming methods *e.g.* Data envelope analysis (DEA) and Benefit of doubt (BoD),
- Budget allocation (BAL),
- Public opinion,
- Conjoint Analysis, and
- Analytical Hierachy Process (AHP).

Nardo *et al.* (2005)

In the field of integrated sustainability assessment, it is important to ensure trade-offs are widely understood and transparent (Section 2.1.3). Weighting processes allow trade-offs between SI_i to be made explicit. Nardo *et al.* (2005:74) provide a tabular summary of the advantages and disadvantages of different weighting methods. Table 2 provides a distillation of that summary in the context of the UPSW project.

Linear programming methods, for example, data envelope analysis (DEA), are beyond the scope of the UPSW project in its proof of concept and pilot phases, as these methods require specification of a sustainability frontier against which baseline states and urban development options may be contrasted. In further development of the UPSW tool, the identification and estimation of sustainability frontiers may prove a useful sophistication.

Statistical methods such as PCA are not applicable because the data for the CI are generated from a network knowledge model. The data requirements for PCA exceed data created by the Bn approach. In the absence of PCA, substitute techniques may be able to assess relationships between SI_i that may compromise the efficiency and consistency of the composite index –

redundancy, causal relationships, and pair-wise correlation. For example, an expert panel or literature searches may fill that role.

Expert opinion methods increase the legitimacy of the composite index. A mix of technical and qualitative input is possible depending on the constitution of the panel. The method is complementary to a knowledge network approach to data generation.

Table 2. Assessment of weighting schemes.

Weight derivation process	Advantages	Dis-advantages
Principal components analysis	Statistical accuracy. Mitigates the double counting problem.	Data requirements not suited to Bn data generation. All SI must have same unit of measurement.
Data envelope analysis (*e.g.* Benefit of doubt)	Clear depiction of progress toward target states.	Requires construction of a sustainability frontier against which options are assessed. Data intensive. Not applicable to UPSW context.
Budget allocation (Expert panel approach)	Uses an expert panel: contribution from diverse fields is possible.	Requires assembly of expert panel. Potentially numerous iterations of the budget (weight) allocation process are required to achieve convergence across the panel.
Public opinion	Focus on the notion of concern or importance to the public.	Not technically consistent with UPSW.
Conjoint Analysis	Allows assessment by disaggregating preferences. Allows willingness to pay assessment. Can be used as an aggregation scheme.	Highly sophisticated statistical techniques. Data requirements may be large. Skepticism as to role as a valuation tool. Implies compensability.
Analytical Hierarchy Process	Technically rigorous. Can be used as the numeric basis for combined schemes such as public opinion and expert panels.	Some criticism exists that delivery of consistent ranking of options in large option sets may not be possible.

Public opinion polls allow wide stakeholder participation. For ease of implementation they may be conducted over large samples (*circa* 60) representative of the principal themes of community concerns. The use of stakeholder panels to design and populate the information requirements for knowledge network models alleviates inconsistency between the qualitative nature of public opinion and the technical/mathematical nature of the UPSW software environment.

Conjoint analysis, specifically choice modeling, has been employed internationally to estimate the non-market value of freshwater systems (Boyle *et al.* 1999; Braden and Johnston 2004; Morrison and Bennett 2004; Hanley *et al.* 2006). The technique has been applied in the jurisdictions of both Auckland City (Kerr and Sharp 2008; Batstone and Sinner 2009) and Canterbury region (Kerr and Greer 2004). As a modeling technique for generating weights for aggregation, it has limitations in that it requires specification of a utility function, the outcomes depend on the sample over which data collection is made, and the estimation process is complex.

4.6.1. The Analytical Hierarchy Process

This review identifies the Analytical Hierarchy Process (AHP) as the preferred method for identifying weights on the sub-indicators that contribute to composite indicators. This approach to creating a weighting system has advantages for the UPSW project in the pilot phase in that it –

- Provides a weighting technique that is consistent with data generated by knowledge network models,
- Considers pair-wise trade-offs between SI_i ,
- Has a sound theoretical basis in the mathematics of linear algebra and is widely used in multi-criteria analysis, providing adequate precedent for its adoption, and
- Has the capacity to integrate a number of techniques by providing a transparent framework that elicits preferences from experts, decision makers, and stakeholders, making trade-offs explicit.

AHP was developed as a general theory of measurement. It is has found wide application in multi-criteria decision making, planning, and decision making. In its general form it is a process for carrying out both inductive and deductive thinking by taking several factors into consideration at the same time, making trade-offs in a numerical format (Saaty 1987).

There are issues around the notion of compensability that are addressed in Section 4.7.1. AHP is a compensatory technique that allows strengths in one sub-indicator to compensate for weaknesses in another. This has implications for the sensitivity of the resulting composite index to changes in the sub-indicators as they reflect changes to the stressors associated with urban development (Nardo *et al.* 2005).

The knowledge network model forms a hierarchy that models the underlying problem, terminating with a set of attributes (sub-indicators). AHP uses a semantic scale over the interval 1 to 9 to assess preferences elicited from pair-wise comparisons between the sub-indicators that contribute information to the composite. A preference of 1 indicates equality between $SI_{i.}$, while a preference of 9 indicates that one SI_i is 9 times more important than another SI_i . Table 3 describes Saaty's (1987) fundamental importance scale.

Table 3. The AHP scale of importance.

Intensity of importance on an absolute scale	Definition	Explanation
1	Equal importance.	Two sub-indicators contribute equally to the composite indicator.
3	Moderate importance of one over another.	Experience and judgment strongly favour one sub-indicator over the other.
5	Essential or strong importance.	
7	Very strong importance.	A sub-indicator is strongly favoured over another, its dominance is demonstrated in the literature or in experience.
9	Extreme importance.	The dominance of one sub-indicator over another is of the highest possible order of affirmation.
2,4,6,8	Intermediate values between two adjacent judgments.	Affords the full range of choices at integer level sensitivity over the range [1 … 9]

Source: Saaty (1987:163).

4.6.2. The pair-wise comparison process: the comparison matrix

Each sub-indicator is compared with each of the other sub-indicators. The outcomes populate a symmetrical matrix in which the lower off diagonal elements of the matrix are reciprocals of the corresponding upper off diagonal elements of the comparison matrix. Depending on the way the DSS is being used, the values assigned to the weights may be generated by decision makers or from the wider community of stakeholders.

Consider the outcomes of a knowledge network model in terms of three sub-indicators $SI_{i.}$. The SI_i are labeled $i = j, k, m$. The corresponding importance weights (w_i) are labeled correspondingly. Table 4 describes the comparison matrix that arises from the pair-wise comparison process.

Table 4. Comparison matrix example.

	$\mathbf{IS_j}$	$\mathbf{IS_k}$	$\mathbf{IS_m}$
$\mathbf{IS_j}$	1	IS_k / IS_j	IS_m / IS_j
$\mathbf{IS_k}$	IS_j / IS_k	1	IS_m / IS_j
$\mathbf{IS_m}$	IS_j / IS_m	IS_k / IS_m	1

In the language of mathematics, the weights of the sub-indicator set are calculated using an eigenvector process in which the eigenvalues of the comparison matrix represent the relative weights, and the rankings of the sub-indicators may be inferred from the relative magnitudes of the weights (Nardo *et al.* 2005).

4.7. Aggregation of component indicators

This section of the report explores the issue of aggregation of sub-indicators to form combined indicators. In the context of the UPSW project, two aggregation schemes have relevance: additive and geometric. The key difference between the two is that the additive scheme implies full compensability, while the geometric option limits compensability.

The type of aggregation scheme employed is strongly related to the normalisation method enacted on raw scores that emerge from the knowledge network process. Linear additive aggregation yields meaningful composite indicators only where sub-indicators are expressed on fully comparable scales. Where this property is not present geometric aggregation should be used.

Arrow’s impossibility theorem (Arrow 1963) demonstrates that no perfect aggregation scheme can exist. It is important that the key properties of the sub-indicators in respect of the issues the composite is assessing are not lost in the aggregation process. Sensitivity analysis should be undertaken to investigate the degree to which the final composite indicators respond to the differing aggregation options.

4.7.1. Compensability

The appropriate normalisation method that precedes aggregation to a combined indicator has been described in Section 4.5. However the issue of compensability remains influential in the choice of aggregation schemes.

Compensability is the capacity for poor performance of one or more sub-indicator to be offset by strong performance in others. Linear aggregation schemes imply full compensability, while geometric aggregation limits compensability. In additive schemes the weights assigned to sub-indicators have the meaning of substitution rates, *i.e.* the trade-offs between sub-indicators. When full non-compensability is required a non-compensatory multiple criteria approach (Nardo *et al.* 2005:76) is appropriate. Section 4.7.4 further discusses the issue of arithmetic versus geometric aggregation schemes.

4.7.2. Additive aggregation

An additive aggregation scheme uses a linear additive model to combine the sub-indicators with their corresponding weights to create a composite indicator.

$$CI = \sum_{1}^{n} w_i * SI_i$$ Equation (4)

Where,

$$\sum_i w_i = 1 \text{ and } 0 \leq w_i \leq 1 \text{ for all } i = 1 \ldots n$$

4.7.3. Geometric aggregation

A geometric weighting scheme uses a multiplicative process to combine the sub-indicators with their corresponding weights to create a composite indicator.

$$CI = \prod_{i=1}^{n} SI_i^{w_i}$$ Equation (5)

Where,

$$\sum_i w_i = 1 \text{ and } 0 \leq w_i \leq 1 \text{ for all } i = 1 \ldots n$$

An alternate formulation of the geometric aggregation process is to use a weighted geometric mean of the form:

$$CI = \left(\prod_{i=1}^{n} SI_i^{w_i} \right)^{1/\sum_{i=1}^{n} w^i}$$ Equation (6)

4.7.4. Additive versus geometric aggregation schemes

Geometric aggregation may be favoured since the arithmetic mean approach may prove vulnerable to either under- or overstatement of the combined index in the presence of outlier values. Freshwater ecologists regard geometric aggregation as preferable in assessing in-stream data since the geometric mean process is more resistant to outlier effects that the arithmetic mean approach. The choice should be made based on sensitivity analysis. Both should be calculated to assess how the results would change. Nardo *et al.* (2005) specify consistency of rankings between development options as the key criteria for selection of an aggregation scheme.

4.8. Sensitivity analysis: the combination of sensitivity and uncertainty analyses (SUA)

The approach to sensitivity analysis reported in this document is based on recommendations of Nardo *et al*. (2005). Those authors advocate the combination of sensitivity and uncertainty analyses (Saisana *et al.* 2005) to mitigate the problem of information loss through aggregation alluded to in Section 4.7.4. This approach allows understanding of the sensitivity of composite indicators to their components and the mathematical processes that created them.

The composite indicator development process involves steps where subjective judgments need to be made: the selection of sub-indicators, the normalisation method, the choice of aggregation model, and the weights assigned to sub-indicators. Sensitivity/uncertainty analysis (SUA) allows assessment of how variation in the outputs generated by the composite process can be apportioned to variation in the underlying assumptions and subjective judgments that have been made.

Criteria for model evaluation will be case dependant, although in general the objective is consistency of rankings and minimisation of average rank shifts across alternative components and combination processes. This information not only informs model development, but allows DSS users to understand the sensitivity of development option rankings to component influences. SUA may also be a forum to make the operation of the composite indicator transparent to stakeholders and decision makers, and to form the basis for discussion of options that emerge from use of the indicator tool.

4.8.1. General framework

The UPSW DSS appraises the effects on receiving water bodies that arise from alternate development options, characterised as the baseline development option (BDO), and a number of alternate urban development options (UDO). In each implementation the composite indices associated with each UDO (CI_{UDO}) are generated by the DSS process. Discrimination between the UDOs is on the basis of the relative rankings based in the magnitude of the $CI_{UDO.}$

A number of uncertainties and subjective decisions need to be evaluated. The quality of the model that has generated the composite indices may be tested by evaluating how the composite indicator varies with changes to the following:

- The contributing set of sub-indicators (subjective decision),
- The normalisation method (subjective decision),
- The distribution of weights created through the AHP process (Section 4.6.1.) (uncertainty), and
- The choice of aggregation scheme (subjective decision).

If CI_{UDO} is the index value for each urban development option ($UDO = 1..max$), R is the set of contributing sub-indicators (R = 1 to r_{max}), S is the set of normalisation schemes (S=1,2,3), and J is the set of aggregation schemes (J=1,2), then the value of CI_{UDO} is,

$$CI_{UDO} = f_J(SI_{R,S}; w_{SI}) \qquad \text{Equation (7)}$$

4.8.2. Monte Carlo Simulation

In order to combine the sensitivity and uncertainty analyses it is appropriate to conduct the process analysis as a single Monte Carlo experiment. This approach incorporates all the uncertainty and subjective factors simultaneously and captures any synergies that may exist among these factors.

Advances in computational software make Monte Carlo simulation an efficient method that allows the influence on the model's output statistic(s) of every expected possible value for each input parameter to be assessed in probabilistic terms. It is similar to deterministic scenario based analysis in that it creates a large number of scenarios. Each specific combination of CIUDO is a "scenario".

In Monte Carlo simulation, in each scenario the value of each input parameter (UDO, R, S,J) that enters the model is randomly selected on a probability weighted basis. For each scenario combination the model is calculated, the output value (CIUDO) stored, and the distribution of output values displayed both as a range and in probability weighted terms (Vose, 2001).

Outputs

Two outputs may be generated by the sensitivity/uncertainty analysis (Nardo *et al.* 2005).

- The rank of each UDO, defined as $Rank(CI_{UDO})$, and
- The average shift ($\bar{R}$) in each UDO ranking for selected set of input parameters over the iterations of the simulation.

Rank (CI_{UDO}) is calculated for each UDO in each scenario on the basis of the relative numerical values generated in the model for each $CI_{UDO.}$

$\bar{R}$ is calculated as:

$$\bar{R} \quad = \quad \frac{1}{M}\sum_{1}^{M} |Rank_{ref}(CI_{UDO}) - Rank(CI_{UDO})| \qquad \text{Equation (8)}$$

A base case is chosen as the reference case for comparison from amongst the options for the possible sets of contributing sub-indicators (*R*), the normalisation schemes, *(S)*, and the aggregations schemes (*J).*

Exploration of *Rank* (CI_{UDO}) and $\bar{R}$ is the extent of the scope of the uncertainty and sensitivity analysis. The model may be used on a case by case basis to explore other issues as required. For example, the Monte Carlo software package @RISK (Palisade Software) has a feature that calculates and displays the correlations between the input and output variables, effectively identifying the contribution of each input variable to total the variance in the output variables.

Inputs

Inputs to the analysis are of two kinds. First the weights, which are subject to uncertainty generated by the subjective outlook of those creating the weights in an AHP process. Second, those aspects of the composite process that are the outcome of a subjective choice by the DSS developers. In this application they are *R*, *S*, and *J*.

The AHP process allows a probability distribution (pdf) of weights for each SI to be obtained over the sample from which the weights are derived. This may be from a group of decision makers, or a wider group of stakeholders, for example. In each scenario the weight WSI is randomly selected on a probability weighted basis from this pdf. The process is repeated iteratively a large number of times (10,000 may be an appropriate number of iterations)

R, S, and J are defined over finite alternatives as discrete random variables. Selection amongst the options for these input factors is achieved by iteratively drawing a random number β, uniformly distributed between [0,1] and applying the Russian roulette algorithm (Nardo *et al.* 2005).

Consider the aggregation scheme input variable J, which may take one of two formats: linear aggregation (J1) or geometric aggregation (J2). For each draw of β, if $\beta \in [0,0.5]$ select J1. If $\beta \in [0.5,1.0]$ select J2. For that iteration CIUDO is calculated using the aggregation scheme alternative triggered by the resultant value of β. This process is followed until a full set of input variables and processes is created sufficient to compute Rank (CIUDO) and $\bar{R}$.

5. CONCLUSIONS

This document has developed recommendations for a sustainability composite indicator. The foundation of the method is based in a review of best practice for integrated sustainability assessment from the perspective of governance and management. The method draws heavily on the work of the OECD Statistics Directorate, with their highly techno-centric approach. The method presented here modifies the OECD approach by allowing non-expert and non-technical specification of weights akin to the use of elements of multi-criteria analysis to address themes that emerged from the governance review.

Those themes include wide stakeholder engagement beyond local and regional government stakeholders to unorganized citizenry impacted by the use of the DSS tool.

Knowledge network models are recommended for data generation because they are multi-functional. They serve as a conceptual map for the urban storm water problem, identifying linkages and dealing with the uncertainty between cause and effect in the stressor-outcome process. Stakeholder and decision maker involvement is a key feature of the method.

Normalisation processes convert raw scores that emerge from final nodes of a Bayesian Network Model to sub-indicators on common interval scales that are consistent with the use of a linear additive model for aggregation. The normalisation process allows options for sub-indicator performance against targets and limits to be assessed.

Weights are generated through the Analytical Hierarchy Process. AHP is a transparent procedure that can be used for both quantitative and qualitative data. It combines theoretical robustness with the flexibility to allow experts, stakeholders and decision makers to assess the trade-offs between sub-indicators through a paired comparison process.

Alternatives for aggregation are offered in the form of linear and geometric aggregation schemes. While the linear approach affords full compensability, the geometric alternative limits the degree that strong performance of one sub-indicator compensates for poor performance of another in the final composite indicator score.

Sensitivity analysis is a key final step in the methodology. Each step of the process that maps raw scores on key variables to the composite indicator is the subject of judgment calls. The way the resulting composite indicator responds to differing input information and alternative approaches to normalisation, and aggregation schemes along with uncertainty associated with sub-indicator weights should be tested to establish the robustness of the composite indicator.

Development of the sustainability indicator system should proceed in two phases.

In the first phase the scope of the indicators should be limited to the estuarine and freshwater receiving bodies, with the DSS tool focused on providing planners and technical experts with depictions of the effects of alternate urban development scenarios. The boundary of the system in question would be defined in terms of the through-flow channels through which the water cycle proceeds from rainfall to freshwater and estuarine receiving bodies. This definition effectively excludes terrestrial effects, such as costs and benefits of alternate urban development options that are not associated with receiving water bodies.

The second phase DSS builds on phase one in two ways. First by expanding the system definition to include the terrestrial urban effects and secondly, by configuration for use beyond experts and decision makers. The tool should be expanded to encompass the entire social-ecological system, moving from a system boundary which excludes terrestrial effects. Further, the DSS configuration should be modified to generate the capacity for its use to be expanded into involvement roles in adaptive governance and social learning processes. The resilience concept provides an avenue to expand the scope of the DSS tool beyond the ecological system to encompass both terrestrial urban, social and ecological systems. It provides a framework with a reasonable pedigree to move from system definition focused on the biophysical receiving water bodies to a whole system focus which spans the entire social-ecological system of human urban development and associated catchment and downstream effects.

6. REFERENCES

Arrow KJ 1963, Social choice and individual values, 2nnd edition, Wiley New York.

Baines JT, Morgan B 2006. Review of Integrated Impact Assessment methods and tools. Prepared for the Ecologic Foundation, with funding from the New Zealand Foundation for Research, Science and Technology. April 2006. 68p.

Batstone C, Sinner J 2009. Techniques for Evaluating Community Preferences for Managing Coastal Ecosystems. Auckland Region Storm Water Case Study. Discrete Choice Model Estimation. Prepared for Auckland Regional Council. Cawthron Report No. 1688. 33 p.

Bebbington J, Brown J, Frame B 2007. Accounting technologies and sustainability assessment models. Ecological Economics, 61(2-3): 224-236.

Beyeler C 2008. DayWater ADSS: an incentive for action but a source of questions - an end-user account of the prototype development. Chapter 16 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management.IWA Publishing, UK.

Boyle M A, Poor J, Taylor L O 1999. Estimating the Demand for Protecting Freshwater Lakes from Eutrophication American Journal of Agricultural Economics 81 1118-1122.

Braden J B, Johnston D M 2004. Downstream Economic Benefits from Storm-Water Management Journal of Water Resources Planning and Management 130 498-505.

Brinsmead T 2005. Integrated Sustainability Assessment: Identifying Methodological Options. Report prepared for the Joint Academies Committee on Sustainability. http://www.naf-forum.org.au/papers/Methodology-Brinsmead.pdf

Carpenter S, Walker B, Anderies JM, Abel N 2001, From Metaphor to Measurement: Resiliience of What to What? Ecosystems 4: 765-781.

Carpenter SR *et al.* 2009. Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment. Proceedings of the National Academy of Sciences, 106(5): 1305-1312.

CSIRO Sustainable Ecosystems 2007. Multi Criteria Analysis Tool (MCAT) User Guide (beta version), St Lucia Qld, Australia, 2007, 29 pages.

Deutsch JC, Metelka T 2008. Adaptive DSS - a new kind of DSS. Chapter 2 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management. IWA Publishing, UK.

Dobes L, Bennett J 2009. Multi-Criteria Analysis: "Good Enough" for Government Work? In Forum: Is Australia's Good Governance in Jeopardy, Agenda, Volume 16, Number 3, available at http://epress.anu.edu.au/agenda/016/03/mobile_devices/ch02.html

Elkington S 1999. Cannibals with Forks: The Triple Bottom Line of 21st Century Business, Capstone Publishing Second Edition, Oxford. Not seen, cited in CSIRO Sustainable Ecosystems (2007).

Ellis JB, Revitt DM, Scholes L 2008. The DayWater Multi-Criteria Comparator. Chapter 9 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management.IWA Publishing, UK.

Figueira J, Salvatore G, Ehrgott M (eds) Multiple criteria decision analysis: state of the art surveys. Springer, Berlin Heidelberg New York.

Folke C 2006, Resilience: the emergence of a perspective for soicial-ecological systems analyses, Global Environmental Change 16(3), 253-267.

Gibson RB 2006. Sustainability assessment: basic components of a practical approach. In Impact Assessment and Project Appraisal, vol.24, no.3, pp.170-183.

Hajkowicz S, Collins K 2007. A review of Multiple Criteria Analysis for Water Resource Planning and Management. Water Resource Management 21: 1553-1566.

Hanley N, Wright R E, Alvarez-Farizo B 2006. Estimating the economics value of improvements in river ecology using choice experiments: and application to the water framework directive Journal of Environment Management 78 183-193.

Holman IP, Rounsevell MDA, Cojacaru G, Shackley S, McLachlan C, Audsley E, Berry PM, Fontaine C, Harrison PA, Henriques C, Mokrech M, Nicholls RJ, Pearn KR, Richards JA 2008. The concepts and development of a participatory regional integrated assessment tool. Climatic Change 90:5–30. DOI 10.1007/s10584-008-9453

Hume T Morrisey D Elliott S 2008, A knowledge network model to inform sediment management_ Mahurangi proof of concept (phase 1) NIWA client report: HAM200-39

Kerr G N, Greer G 2004. New Zealand River Management: Economic Values of Rangitata River Fishery Protection Australian Journal of Environmental Management 11 139-149.

Kerr GN, Sharp BMH 2004. Evaluating off-site environmental mitigation using choice modelling, The Australian Journal of Agricultural and resource Economics, 52, pp 381-389.

Leitcher R, Wiederman S 2004. Modelling Economic-Ecological Impacts in NSW catchments_Feasibilty Study, Integrated Catchment Assessment Centre (iCAM) Client Report to NSW Department of Environment and Conservation, The Australian National University, Canberra.

Maani KE , Cavana RJ 2000. Systems Thinking and Modelling: Understanding Change and Complexity, Pearson Education Auckland NZ 262p.

Morriso, M, Bennett J 2004. Valuing New South Wales rivers for use in benefit transfer The Australian Journal of Agricultural and Resource Economics 48 591-611.

Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovannini E 2005. Handbook on Constructing Composite Indicators: Methodology and User Guide. Directorate OS, ed. OECD OLIS, Paris.

Parsons M, Thoms M, Capon T, Capon S, Reid M 2009. Resilience and thresholds in river ecosystems, Waterlines Report Series No21, September 2009, Australian Government National Water Commission, p97.

Perez P, 2010 pers. comm. 17 June 2010.

Proctor W, Qureshi ME 2005. ‘Multi-criteria Evaluation Revisited’, Proceedings of the Australia New Zealand Society for Ecological Economics Conference, Massey University, Palmerston North, New Zealand, 11-13 December, NZCEE.

Proctor W, Drechsler M 2003. Deliberative Multi-criteria Evaluation - A case study of recreation and tourism options in Victoria, Australia in: 57th Meeting of the European Working Group Multiple Criteria Decision.

Quinn JM Monaghan RM Wilcock RJ 2009. Linking Farm and Waterway Values - The Bog Burn Catchment, Conference presentation, personal communication John Quinn.

Quinn J 2011. Hurunui BBN Assumptions, Personal Document John Quinn NIWA, February 2011.

Saaty RW 1987, The Analytical Hierachy Process- What It Is And How It Is Used, Mathematical Modelling, Vol 9 No 3-5, pp 161-176.

Sadler B, Ward M, Frame R 2008. A Framework for Sustainability Appraisal in New Zealand - Thinkpiece. Landcare Research Contract Report LC0708/090.

Saisana M, Nardo M, Satelli A 2005. Uncertainty and Sensitivity Analysis of the 2005 Environmental Sustainability Index, in Esty D, Levy M, Srebotnjak T Sherbinin A. 2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship. New Haven: yale Centre for Environmental Law and Policy, p 75 – 78.

Taylor A 2005. Guidelines for Evaluating the Financial, Ecological and Social Aspects of Urban Stormwater Management Measures to Improve Waterway Health, Cooperative Research Centre for Catchment Hydrology, Technical Report Report 05/11, 168 pp.

Thevenot DR *et al.* 2008. DayWater: the challenge to develop an Adaptive Decision Support System for urban stormwater source control. Chapter 1 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management.IWA Publishing, UK.

Ticehurst J L, Newham L T H, Rissik D, Letcher R A, Jakeman A J 2006. A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia. Environmental Modelling & Software 22 pp 1129-1139.

van den Belt M O, Bianciotto R, Costanza S, Demers S, Diaz G, Ferryra E, Koch F Momo, M Vernet 2006. Mediated Modeling of the impacts of Enhanced UV-B Radiation on Ecosystem Services. Special Issue of Photochemistry and Photobiology, 82:865-877.).

van den Belt M 2004. Mediated Modeling; a system dynamics approach to environmental consensus building. Island Press, Washington D.C.

Videira N, Antunes P, Santos R, Lopes R 2010. A Participatory Modelling Approach to Support Integrated Sustainability Assessment Processes. In Systems Research and Behavioural Science Vol.27, 446-460.

Vose D 2001. Risk Analysis, A Quantitative Guide, 2nd edition, Wiley, UK.

Walker B, Salt D 2006. Resilience thinking: sustaining ecosystems and people in a changing world: how can landscapes and communities absorb disturbance and maintain function? Island Press, London, 174p.

Walmsley J 2002. Framework for measuring sustainable development in catchment systems. Environmental Management 29 (2), 195–206.

Ward TJ, Butler A 2006. 'Coasts and oceans', theme commentary prepared for the 2006 Australia State of the Environment Committee, Department of Environment and Heritage, Canberra, available at url:http://www.deh.gov.au/soe/2006/commentaries/coasts/index.html

Websites

Farr DF, Rossman AY, Palm ME, McCray EB undated. Fungal Databases, systematic botany & Mycology Laboratory, ARS, USDA. Retrieved 6 April 2003, from http://nt.ars-grin.gov/fungaldatabases

7. APPENDICES

Appendix 1. Bibliography of additional sources reviewed in preparation of this report.

Cork S, Stoneham G, Lowe K 2007. Ecosystem Services and Australian Natural Resource Management (NRM) Futures. Paper to the Natural Resource Policies and Programs Committee (NRPPC) and the Natural Resource Management Standing Committee (MRMSC).

Dahl A L 2000. Using Indicators to Measure Sustainability: Recent Methodological and Conceptual Developments. Marine Freshwater Research (51) pp 427-433.

Davies S, Jackson S K 2006. The Biological Condition Gradient: A Descriptive Model for Interpreting Change in Aquatic Ecosystems. Ecological Applications by the Ecological Society of America. 16 (4) pp. 1251-1266.

Eigenbrod F, Armsworth P R, Anderson B J, Heinemeyer A, Gillings S, Roy D B, Thomas C D, Gaston K J 2010. The impact of proxy-based methods on mapping the distribution of ecosystem services. British Ecological Society - Journal of Applied Ecology 47 pp 377-385.

Gal G, W Anderson 2009. Methods in Ecology and Evolution: A novel approach to detecting a regime shift in a lake ecosystem. British Ecological Society.

Gomez-Limon J A, Risego L 2009. Alternative approaches to the construction of a composite indicator of agricultural sustainability: An application to irrigated agriculture in the Duero basin in Spain. Journal of Environmental Management 90 pp 3345 – 3362. www.elsevier.com/locate/jenvman

Gomez-Limon J A, Sanchez-Fernandez G S 2010. Empirical evaluation of agricultural sustainability using composite indicators. Ecological Economics 69 pp 1062 – 1075. www. elsevier.com/locate/ecolecon

Hammond A *et al.* 1995. Environmental Indicators: A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development. World Resources Institute.

Johnston R, Besedin E Y, Ranson M H 2006. Characterizing the effects of valuation methodology in function-based benefits transfer. Ecological Economics 60 pp 407-419. www.elsevier.com/locate/ecolecon

Kay J *et al.* 1999. An Ecosystem Approach for Sustainability: Addressing the Challenge of Complexity. Futures (31) 7 pp 721-742.

McEvoy D, Ravetz J 2001. Impact Assessment and Project Appraisal: Toolkits for regional sustainable development 19 (2) pp 90-93.

Milman A, Short A 2008. Incorporating resilience into sustainability indicators: An example for the urban water sector. Global Environmental Change 18 pp 758-767. www.elsevier.com/locate/gloenvcha

Nijkamp P, Ouwersloot H. A Decision Support System for Regional Sustainable Development: The Flag Model.

Noble B F 2000. Environmental Impact Assessment Review: Strengthening EIA through adaptive management: a systems perspective 20 pp 97-111.

Olsson P, Gunderson L H, Carpenter S R, Ryan P, Lebel L, Folke C, Holling CS 2006. Shooting the Rapids: Navigating Transitions to Adaptive Governance of Social-Ecological Systems. Ecology and Society 11 (1): 18.

Ravetz J 1999. Environmental Impact Assessment Review: Integrated assessment for sustainability appraisal in cities and regions (20) pp 31-64.

Ravetz J 2000. Environmental Impact Assessment Review 20. pp31-64.

Rees W E 2003. Understanding Urban Ecosystems: An Ecological Economics Perspective pp1-18.

Robson B J, Webster I T, Hamilton D P, Chan T. Ten Steps Applied to Development and Evaluation of Process-Based Biogeochemical Models of Estuaries.

Sinner J, Baines J, Crengle H, Salmon G, Fenemor A, Tipa G 2005. Key concepts in Sustainable Development Part 2: The social dimension. Public Sector 28 (4).

Sinner J, Baines J, Crengle H, Salmon G, Fenemor A, Tipa G 2005. Key concepts on Sustainable Development Part 1: Economics and Ecology. Public Sector 28 (3).

Sustainability Assessment Symposium 2010. Towards Strategic Assessment for Sustainability. Integral Sustainability.

Ticehurst J L, Newham L T H, Rissik D, Letcher R A, Jakeman A J 2006. A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia. Environmental Modelling & Software 22 pp 1129-1139.

Walker BH, Gunderson LH, Kinzig AP, Folke C, Carpenter S R, Schultz L 2006. A handful of heuristics and some propositions for understanding resilience in social-ecological systems. Ecology and Society 11(1): 13. [online] URL: http://www.ecologyandsociety.org/vol11/iss1/art13/

Walmsley J J 2002. Framework for Measuring Sustainable Development in Catchment Systems. Environmental Management 29 (2): pp195-206.

Appendix 2. Sustainability Assessment - Governance and Management Issues: Literature Review Report.

Sustainability Assessment - Governance and Management Issues: Literature Review Report

Prepared by James Baines and Bronwyn Morgan
Taylor Baines & Associates

for the Cawthron Institute

October 2010

Table of Contents

EXECUTIVE SUMMARY

Contract brief

1. This document reports on a desk-based search, retrieval and review of literature on sustainability and integrated assessment to produce recommendations for best practice as understood at this time for use as a guide to Cawthron's process development work.

2. The report will evaluate requirements for effective sustainability assessment, including governance and management issues, and will:

   1) identify possible approaches to implement an integrated sustainability indicator system for urban development scenario assessment; and

   2) make recommendations from discoveries in (1) based on a SWOT analysis for an approach to further develop and implement the proof of concept currently under investigation for the DSS

Part 1: Requirements for effective sustainability assessment

3. The literature canvassed in this review has drawn from experience in North America, Europe and Australasia. Considerable cross-fertilisation of ideas and initiatives is evident from the bibliographies, although clustering into distinctly separate citation sequences is also evident.

4. The literature search encountered a number of emerging and over-lapping fields of enquiry, experimentation and practice. For simplicity, the term Sustainability Assessment will be used hereafter to include such variants as Sustainability Appraisal, Integrated Assessment or Integrated Sustainability Assessment.

5. Formal applications of Sustainability Assessment encountered so far are predominantly at the strategic level. That is to say, they are generally aimed at public policy decisions or strategic public agency decisions.

6. Analytical frameworks for Sustainability Assessment are still generally at the developmental or trial stage. However, these same characteristics - strategic level and developmental stage - are precisely those which apply to the research context for this literature review.

7. In 2006, Brinsmead's work was the only literature which identified at a fundamental conceptual level what integrative functions are necessary for the practice of integrated assessment. This review update has been unable to re-contact Brinsmead; nor has it found any further elaborations or extensions of his work; nor has it found any other comparable analysis or first principles-based review of the practice of 'integration' in assessment activities. However various authors (Gibson (2006), Videira et al. (2010), Sadler (2008)) have attempted to make progress in articulating fundamental principles of good Sustainability Assessment practice.

8. Two main themes have emerged from this update on integrated assessment. The first theme is the increasing recent focus on Sustainability Assessment as a structured

assessment methodology, as distinct from the earlier more general theme of integration in assessment. Within this methodological focus, sub-themes include -

- perspective on sustainability;
- identifying objectives for assessment and decision making;
- clarifying trade-off rules;
- adaptive processes; and
- relationship between those involved in assessing and those involved in decision making.

9. The second main theme is stakeholder engagement as a specific focus of practice, exhibiting a maturing interest and more sophisticated discourse about rationales and approaches to stakeholder engagement in Sustainability Assessment processes. Within this focus, the sub-themes are -

- greater differentiation and specification of stakeholder roles and functions, and
- the phenomenon of social learning.

Perspective on sustainability - characterising the current position

10. Several writers with longstanding involvement in the development of assessment tools preface their discussion of practical Sustainability Assessment methodologies by conceptualising the pursuit of sustainable development primarily in terms of remedying what can clearly be seen already to be unsustainable and at risk, rather than aiming for ill-defined and difficult to agree aspirations.

11. Their framing of the sustainability challenges is similar to the framing of the NIWA/Cawthron experimental concept, with its identification of impacts on receiving waters (environmental capital) of current stormwater management practices and the reflection of these impacts in ecological, social and financial values. In short, decision making aids such as Sustainability Assessment should incorporate explicit consideration of limits and targets as they relate to the management of capital stocks, and most likely need to recognise that some of these limits may already have been transgressed in particular locations or may be at risk of being transgressed.

Identifying objectives to be applied in the assessment - sustainability concept definition

12. The step change from earlier versions of assessment practice is the shift from relative assessments which focus on assessing whether or not proposals or policies result in outcomes which are generally in the desired direction (“direction to target” assessments) to absolute or comparative assessments which focus on assessing how far a particular proposal or policy will steer outcomes towards desired targets (“distance to target” assessments).

13. For such assessment procedures, objectives have to be defined in a manner which renders them amenable to quantification or qualitative comparison.

14. With regard to practical methodologies for assessment and decision making, the central debate here is between those who adopt a 3-pillar approach to defining objectives and associated criteria and indicators, and those who attempt to define sustainability objectives in terms of the observed ‘requirements for improvement’ rather than the established categories of expertise.

15. In the former approach, integration is achieved when economic, environmental and social factors are addressed simultaneously, and evaluated against a sustainability framework derived from international or national policy or strategies. In the latter approach, integration is achieved by avoiding traditional fragmentation and avoiding presumptions of responsibility that come from fragmentation."

16. The NIWA/Cawthron research, involving choice modelling, could be interpreted as moving beyond the pillar approach by focussing on "requirements for improvement" in terms that are recognisable to members of the public - underfoot conditions, water clarity and ecological health. It also encapsulates elements of the 3-pillar approach with its simultaneous consideration of environmental, social and economic factors, and may incorporate the notion of upper and lower thresholds (sustainability targets and limits), depending on how the low/medium/high ranges are specified.

Clarifying trade-off rules - sustainability concept application

17. A common theme in literature on sustainable development is the question of balance in the simultaneous pursuit of multiple objectives. It is difficult to avoid the concept of trade-offs, and the related questions of how trade-offs should be conceived in a sustainability context, and under what circumstances should they be considered.

18. Propositions for trade-off rules tend to adopt a tiered approach to specifying principles, subject to a general requirement not to compromise the achievement of net sustainability gain. They also emphasise requirements for transparency and an obligation on trade-off proponents to provide explicit justification.

19. Without knowing more about the details and assumptions of the choice experiment scenarios envisaged by the NIWA/Cawthron research team, it is not clear what trade-off assumptions are embedded within the overall decision logic. However, good practice will require that any trade-off assumptions and mechanisms that may be <u>implicit</u> within the choice experiments and associated modelling should be made <u>explicit</u> and discussed with the decision makers to ensure that they are aware.

Adaptive process design in Sustainability Assessment

20. Two aspects of 'adaptiveness' were encountered in the literature. The most common interpretation of 'adaptiveness' relates to iterative assessment processes which have evaluation and learning procedures built into them to enable adjustments in decisions over time. The other interpretation of 'adaptiveness' refers to the need to customise assessment processes to the local context.

21. To the extent that the NIWA/Cawthron choice modelling experiment is being applied at three different types of locations - upper harbour, middle harbour, outer harbour - the contextualising/customising mode of adaptiveness may be relevant.

The relationship between assessment and decision making

22. The critical issue here for good practice in Sustainability Assessment is the relationship between the activities of assessment and decision making. Are the decision makers themselves involved directly in the assessment process and thereby intimately familiar with the mix of information and value judgements which have informed the results? The strongest expression of requirement takes the form that

the direct involvement of decision makers in the assessment activities is a central element of integration in the context of Sustainability Assessment.

23. It is not clear from the descriptions received so far of the NIWA/Cawthron choice modelling experiment what relationship is intended between those whose values are expressed in the choice experiment forums and those who will subsequently make decisions using the resulting Adaptive Decision Support System (ADSS).

Greater differentiation and specification of stakeholder roles and functions in integrated assessment

24. The literature reveals considerable evidence of a progressive shift from expert, managerial processes towards deliberative democracy, and provides a variety of explanations for this.

25. This range of justifications points clearly to the potential scope for stakeholder involvement throughout every stage of an integrated assessment process, from issue framing and objective setting, through deliberation and prioritising to decision making.

26. When addressing the topic of stakeholder engagement, an inclusive interpretation of ‘stakeholder’ is advised.

27. This review confirms the importance of stakeholder involvement to the research programme anticipated by NIWA/Cawthron. How the research team has already engaged stakeholders or intends to address questions of stakeholder engagement is clearly important, for all the reasons outlined.

28. The literature makes it clear that initiatives on stakeholder engagement cannot be subject to rigid guidelines but need to be adaptive to the task at hand. Such initiatives inevitably involve an element of experimentation, informed by the accumulating practical experience of facilitated participation processes (that exists within the research and professional community) and a commitment to evaluate effectiveness periodically.

Social learning

29. The literature emphasises the importance of social learning to the contemporary practice of Sustainability Assessment.
30. Certain research tools, such as MCA and mediated modelling, can be helpful in structuring social learning processes.

31. Achieving social learning outcomes amongst participating stakeholders and the NIWA/Cawthron research team will be an important test of the effectiveness of any resulting ADSS.

Part 2: Requirements for implementing an integrated sustainability indicator system (an ADSS based on Multi-Criteria Assessment tool)

32. This review accessed a number of Multi-Criteria Assessment (MCA) cases being applied in the context of Sustainability Assessment and related decision making. Most of these are cases of trialling a general methodology to a particular sustainability challenge.

33. Of the cases reviewed, the DayWater Project is unique in attempting to produce a computer-mediated ADSS tool for remote access by users. All the other cases involve the use of MCA methodologies in interactive group processes, although some of these involved use of computer-based tools mediated by the researchers.

34. The expressed strengths and weaknesses are reviewed to see if they provide additional insights to those already gleaned from the broader Sustainability Assessment literature.

35. The experience of the DayWater Project was reviewed from the reported experiences of the MCA tool developer and a prospective end user to see what additional guidance could be gleaned for the NIWA/Cawthron ADSS development work.

Recommendations

36. **Attention to trade-offs:** good practice will require that any trade-off assumptions and mechanisms that may be implicit within the choice experiments and associated modelling should be made explicit and discussed with the decision makers to ensure that they are aware of the nature of the trade-offs associated with any choice/decision.

37. **Dialogue throughout the development process:** building in both elements of adaptiveness to the ADSS will require close attention to stakeholder/researcher dialogue. In particular, building in evaluation and learning procedures to the methodology to enable adjustments in user decisions over time will require such dialogue to take place (be managed) at various stages of ADSS development, not just at the beginning and end of the development sequence.

38. **Stakeholder selection:** a selection of intended future users/decision makers should be included amongst the stakeholders involved directly and throughout the development of the ADSS. It should not however be assumed that they are the only stakeholders or indeed the most important stakeholders.

39. **Timing of stakeholder engagement:** if stakeholder engagement has not already been initiated, this should happen before the end of the ‘proof of concept’ stage with the expectation of periodic involvement throughout each stage of ADSS development.

40. **Collaboration with other science programmes:** NIWA/Cawthron should consider initiating a science forum devoted to stakeholder engagement issues, with a view to providing a regular forum for exchange of ideas and experience amongst scientists with these common interests.

41. **Expert facilitation:** in order to maximise social learning during the research programme, the NIWA/Cawthron research team should consider the use of expert facilitators to plan and facilitate periodic group work sessions involving stakeholders.

42. **Testing for social learning outcomes:** as a means of monitoring progress toward an effective ADSS, the NIWA/Cawthron team should consider monitoring social learning outcomes amongst participating scientists and stakeholders at intervals during the development process.

INTRODUCTION

1.1 Background to this review

The Cawthron Institute (Cawthron) is participating in a research programme, led by the National Institute for Water and Atmosphere (NIWA), funded by the Foundation for Research, Science and Technology (FRST) and entitled Urban Planning for Sustainable Water Bodies in New Zealand.

Cawthron is responsible for the development of methods to support an integrated sustainability indexing system for expressing the environmental, social, economic and cultural impacts of urban development on water bodies. This indexing system will provide the basis by which alternative development scenarios will be compared by users of a pilot Decision Support System (DSS).

1.2 Contract brief

In February 2010, Cawthron engaged Taylor Baines & Associates (TBA) to provide research assistance.

TBA's original brief was to carry out a desk-based search, retrieval and review of literature on sustainability and integrated assessment and to produce recommendations for best practice as understood at this time for use as a guide to Cawthron's process development work.

Specifically the report will evaluate requirements for effective sustainability assessment, including governance and management issues, and will:

1) identify possible approaches to implement an integrated sustainability indicator system for urban development scenario assessment; and

2) make recommendations from discoveries in (1) based on a SWOT analysis for an approach to further develop and implement the proof of concept currently under investigation for the DSS.

Discussions between Cawthron and TBA during the course of the literature search resulted in some refinement of search and review focus.

1.3 Cawthron-NIWA research focus

The Cawthron-NIWA research programme is prioritised on the following attributes -

1) urban catchments;
2) stormwater and urban run-off; and
3) non-point-source pollution.

1.4 Taylor Baines' experience relevant to this brief

Taylor Baines & Associates was involved in previous FRST-funded research with the Ecologic Foundation - Institutions for Sustainable development, with responsibility for contributions on Social Sustainability and Integrated Impact Assessment methods.

Taylor Baines is currently involved in research on collaborative governance processes with the Ecologic Foundation, with a case study of the Land and Water Forum.

2 WORK PROGRAMME

2.1 Scope of literature search

The literature search had to areas of focus

1) updating the previous review of literature on integrated assessment for guidance on evolving best practice; and

2) review approaches to the use of multi-criteria methods - used when implementing “integrated sustainability indicator systems” - with particular emphasis on governance and management issues.

As for the 2005 literature review, the intended emphasis has been on actual tools, processes or methods for substantive integration, or actual multi-criteria methods used in the context of integrated sustainability indicator DSS.

2.2 Search and retrieval methods

Searching involved computer-based internet and database searching (details are provided in Appendix 5) This was supplemented by email contacts and networks, including other CRI research teams working on related topics, and assistance from the Cawthron librarian with some document retrieval.

2.3 Approach to review and analysis

The literature review was carried out using thematic analysis and keyword summaries.

3 REVIEW FINDINGS

3.1 Two objectives of review

This literature review exercise has had two objectives to focus on. The first objective was to update of the literature review carried out on the topic of Integrated (Impact) Assessment[1] and reported in April 2006. The second objective was to identify possible approaches to implementing an integrated sustainability indicator system for urban development scenario assessment in the context of receiving water bodies for urban storm water.

Arising from these review activities, the report makes recommendations for an approach to further develop the proof of concept for the application of a multi-criteria method to manage urban storm water, involving stakeholder engagement.

3.2 Update on Integrated Assessment

3.2.1 Overview

Geographic origins

The literature canvassed in this review has drawn from experience in North America, Europe and Australasia. Considerable cross-fertilisation of ideas and initiatives is evident from the bibliographies, although clustering into distinctly separate citation sequences is also evident.

Over-lapping fields of enquiry and experimentation

The literature search encountered a number of emerging and over-lapping fields of enquiry, experimentation and practice. These include Integrated Assessment, Sustainability Appraisal, Integrated Sustainability Assessment, Adaptive Management, Adaptive Governance, Collaborative Governance, Participatory Assessment, Mediated Modelling, and so forth.

Such overlapping fields of interest inevitably generate semantic variations. By way of example, many papers which discuss ‘participatory methods’ or ‘stakeholder engagement’ refer to the concept of ‘social learning’. Some papers refer to ‘deliberative democracy’. While the two concepts are clearly not the same, they are also not un-related. Furthermore, at the conceptual level, both concepts have the potential to be of interest in an enquiry about participatory urban planning methodology.

This review has not attempted to characterise and differentiate these over-lapping fields of interest. Instead, the review has maintained a focus on identifying the requirements for effective Sustainability Assessment, drawing from across these fields where cases could be found that provide insights on practical details and evaluation experience. For simplicity, the term Sustainability Assessment will be used hereafter to include such variants as Sustainability Appraisal, Integrated Assessment or Integrated Sustainability Assessment.

The purpose of the observations made in the three foregoing paragraphs is simple. The volume of literature material potentially available and suggesting promise of relevance on

[1]The focus of the 2006 literature review was on integration in impact assessment methodologies; hence the term ‘integrated impact assessment’. In practice, such assessment methodologies have evolved under a variety of headings including Sustainability Assessment, Sustainability Appraisal, Integrated Sustainability Assessment, and so on. The 2006 literature review attempted to be inclusive, irrespective of nomenclature, with a primary focus on integrative practices.

various aspects of integration is immense. The number of published literature items accessed that are of practical value is very limited[2]. Some of the items accessed and reviewed have pointed to other literature items that suggest promise of useful insights[3]. These will be identified for future reference, where appropriate.

Current stage of applications - strategic level, developmental and pilot trials

Formal applications of Sustainability Assessment encountered so far are predominantly at the strategic level. That is to say, they are generally aimed at public policy decisions or strategic public agency decisions. For example, Gibson (2006, p.170) cites the United Kingdom's use in regional planning and Hong Kong's application to evaluating urban infrastructure options. Gibson also notes that *"many private firms have been experimenting with forms of triple bottom-line assessment"*, but later criticises this form of "pillar-based" assessment as representing fragmentation rather than integration.

Analytical frameworks for Sustainability Assessment are still generally at the developmental or trial stage. At the conclusion of their evaluations, Tuinstra et al. (2008, p.149) expressed the hope that *"The findings can also be incorporated into the design of 'real' ISAs, i.e. full-fledged sustainability assessments carried out in a real-world context rather than in a research project."* Sadler (2008, p.5) reflected that *"Methodologically, the emphasis is on exploring 'the art of the possible', bringing together a combination of tested and emerging tools and approaches and following the sage advice of Serageldin (1996, citing Solow) who advocated taking a series of pragmatic steps to improve current approaches incrementally rather than engaging in interminable debate about their imperfections."* . Similarly, Gibson's paper (2006, p.171) *"relies on the past two decades of deliberations and experience with sustainability assessment initiatives, as a foundation for identifying the basic sustainability requirements that should inform a transition to sustainability assessment;"*.

However, these same characteristics - strategic level and developmental stage - are precisely those which apply to the research context for this literature review. This is where the process experimentation has been happening in Sustainability Assessment over the past few years.

3.2.2 Cues from 2006 literature review

Conclusions from the 2006 Review

The 2006 literature review on Integrated Assessment (Baines and Morgan, 2006) did not presume a focus on strategic-level, public-agency applications. Rather it aimed to explore what it means to engage in integrative practice when carrying out an impact assessment of any kind, at any level[4].

The review in 2006 drew the following conclusions -

> (i) the practice of integration in impact assessment is at an experimental, emergent stage, with little documented evidence of conscious reflection and evaluation;

---

[2] As with any literature review exercise, there was a proportion of promising abstracts which yielded little if any practical insights of relevance.

[3] For example, several items reviewed have cited European work by Weaver, P.M. and Rotmans, J. 2006. Integrated sustainability assessment: What, why and how? In *International Journal of Innovation and Sustainable Development*, Vol.1, No.4, pp.284-303. However, despite best efforts by the review team and the Cawthron librarian, this item has yet to be accessed.

[4] i.e. whether at project level, strategy level, plan level or policy level.

(ii) there is an emerging taxonomy of approaches to integrated impact assessment, distinguishing three principal modes of assessment - replicating EIA-based assessment[5], objectives-led appraisal[6], and what was termed principles-based assessment[7];

(iii) the need for methodological pluralism, tempered by consideration of functional equivalence - a variety of methodologies but a common purpose;

(iv) strong themes associated with integrative practice were an emerging recognition of the importance of stakeholder participation, an emphasis on iterative/cumulative assessment procedures and the development of 'a common framework' of understanding the problem situation by those involved in the assessment.

Reflecting on the 2006 conclusions in light of the 2010 Review

Reflecting on these conclusions at the present time, it would appear that -

(i) the formal development of integrative practice is still largely experimental in nature; that is to say, there is little evidence of integrative methods being mainstreamed in policy or strategy development and project evaluation processes. There continues to be little documented evidence of conscious reflection and evaluation in the accessible published literature, further reinforced by Tuinstra et al. (2008, p.130)[8]
(ii) of the three principal modes identified in 2006, pillar-based, objectives-led appraisal appears to have achieved greater prominence, particularly in the form of (Integrated) Sustainability Assessment and Triple Bottom Line applications (Ellis et al., 2008; Lai et al., 2008; Proctor and Drechsler, 2006; Russell and Ward, 2010; Taylor and Fletcher, 2005; Tuinstra et al., 2008)

(iii) the Sustainability Assessment modality reflects the common purpose within which the literature indicates some common elements of process architecture (objective definition, criteria selection, need for trade-off rules, information for decision makers), while also reflecting variation in methodological details.

(iv) the focus on stakeholder participation has attracted much attention, resulting in a more sophisticated discourse about rationales and approaches.

The review in 2006 accessed several much more extensive literature review processes at the time (Dalal-Clayton and Sadler (In Press), Brinsmead (2005), Buselich (2002) and Toth (2003)).

With respect to detailed descriptions of integrating processes that could be replicated or tools that could be tried by others and evaluated, Dalal-Clayton and Sadler concluded *"the discussion of these issues in the impact assessment literature lacks a cutting edge and is largely general or theoretical."* and *"Linking impact assessment to other tools and processes*

---

[5]where the basic one-dimensional mode of assessment is replicated in the so-called 3-pillar form of parallel assessments of environmental, social and economic changes
[6]in which the assessment is carried out within the explicit framework of established policy goals and principles, except that once again it is replicated in the so-called 3-pillar form of parallel assessments
[7]with the attribute of being objectives-led, but where the objectives are derived from broader sustainability principles
[8]Tuinstra et al. observe that *"Documented findings are useful because relatively few assessment processes have been evaluated systematically for learning experiences"*

is also a critical part of an integrated approach to SA (sustainability appraisal)." It is these other tools and processes which have been the focus of more recent developments. Amongst them is the increasing use of multi criteria assessment, various types of modelling for common framework articulation, and increasing refinement of stakeholder-engagement processes.

Update on principles underpinning assessment practice

In 2006, Brinsmead's work was the only literature which identified at a fundamental conceptual level what integrative functions[9] are necessary for the practice of integrated assessment. The review in 2006 (p.9) also noted that *"...many other authors imply a similar taxonomy. The fact that so many authors treat the integrative functions in an implicit manner serves to reinforce the findings described previously ... that much knowledge about the practice of integration resides mainly in the realm of the tacit knowledge of practitioners applying their skills and experience at the workface."*

This review update has been unable to re-contact Brinsmead; nor has it found any further elaborations or extensions of his work; nor has it found any other comparable analysis or first principles-based review of the practice of 'integration' in assessment activities. However various authors (Gibson (2006), Videira et al. (2010), Sadler (2008)) have attempted to make progress in articulating fundamental principles of good Sustainability Assessment practice.

In this regard, there is an important distinction to make between "integrated assessment" and "sustainability assessment". Whereas Sustainability Assessment arguably requires effective integration, an integrated assessment does not have to be in pursuit of sustainable development objectives.

Nevertheless, the 2006 Review findings and their conceptualising of the fundamental integrative functions do in some ways provide a convenient launching point to discuss several recent new directions which have emerged as strong themes in the more recent literature. The maturing of Sustainability Assessment practice per se is one of these new directions, as is the maturing interest and sophistication in stakeholder engagement processes. Both these aspects of innovation enrich the dimension of contextual integration. Furthermore, as will be discussed in more detail in section 3.3.2 below, the growing interest in using multi-criteria analysis as a logical framework within which to structure stakeholder engagement processes has the potential to enrich all the integrative functions.

3.2.3 Themes from recent applications of Integrated Assessment

Two main themes have emerged from this literature review. The first theme is the increasing recent focus on Sustainability Assessment as a structured assessment methodology, as distinct from the earlier more general theme of integration in assessment. Within this methodological focus, various sub-themes suggest issues worthy of additional comment. They include -

- perspective on sustainability;
- identifying objectives for assessment and decision making;
- clarifying trade-off rules;
- adaptive processes; and

[9]The five integrative functions identified by Brinsmead are descriptive integration, evaluative integration, strategic integration, contextual integration and mutual (functional) integration. The last of these refers to the conscious methodological design and management of integrating procedures and behaviours.

- relationship between those involved in assessing and those involved in decision making.

It should be remembered that the recent applications of Sustainability Assessment covered within this review exercise are predominantly (though not exclusively) those involving some form of multi-criteria assessment tool.

The second main theme is stakeholder engagement as a specific focus of practice. Already noted is the maturing interest and more sophisticated discourse about rationales and approaches to stakeholder engagement in Sustainability Assessment processes. Within this focus, several sub-themes are worthy of particular comment. They are -

- greater differentiation and specification of stakeholder roles and functions, and
- the phenomenon of social learning.

3.2.4 Sustainability Assessment

Perspective on sustainability - characterising the current position

Several writers with longstanding involvement[10] in the development of assessment tools (Gibson and Sadler) preface their discussion of practical Sustainability Assessment methodologies with telling statements of philosophy. They conceptualise the pursuit of sustainable development primarily in terms of remedying what can clearly be seen already to be unsustainable and at risk, rather than aiming for ill-defined and difficult to agree aspirations. Gibson (2006, p.171) had this to say -

> *"Attention to sustainability objectives is driven not so much by a desire to preserve tested traditions as by demands for improvements ....the present concept of sustainability is a response to evidence that current conditions and trends are not viable in the long run, and that the reasons for this are as much social and economic as they are biophysical or ecological."*

Their interest is in providing decision makers with tools that will help them improve past decision making practice. Sadler (2008, p.5) put it this way -

> *"Methodologically, the emphasis is on exploring 'the art of the possible', bringing together a combination of tested and emerging tools and approaches and following the sage advice of Serageldin (1996, citing Solow) who advocated taking a series of pragmatic steps to improve current approaches incrementally rather than engaging in interminable debate about their imperfections."*

This challenge for decision making is not so much a retreat from idealism as a determined pursuit of pragmatism; what Gibson describes as *"conceptual rigour and effective action"* in preference to the predominant *"cheerful visions and passionate endorsement"*. Sadler (p.8) observes a similar dichotomy when he states -

> *"Advocates of development take an optimistic stance, speaking of challenges to be overcome in building a more prosperous and secure future, and eschew assessments of the level of threat to sustainability or the likely downside consequences as moot. ....This stance does not so much disavow the notion of limits, rather it considers adjustments must and will be made well before these are reached."*

[10]These authors are among a number who not only have many contributions in their own name, they are commonly associated with others in developmental and evaluation initiatives.

Attitudes to 'limits' are central to the discourse. However, 'limits' mean different things to different actors. To some, limits provide a stimulus to technological or behavioural innovation in order to overcome those limits in the present. To others, limits signal constraints that should not be breached because to do so impacts adversely not just on the present but potentially on future generations.

It is safe to assert that consideration of inter-generational equity features in almost any conceptualisation of sustainable development. Operationalising this concept is a challenge that not many have addressed. The school of practice represented by Sadler and his colleagues around the globe proposes *"The representation of capital stocks as a proxy measure of the opportunities to meet present and future needs"* (Sadler, 2008, p.10). This notion of a positive asset base invites consideration of various components: *"resource potentials"* as environmental or natural capital, *"created wealth"* as financial or economic capital and *"human capabilities"* as social or human capital.

The framing of the sustainability challenges by the likes of Gibson and Sadler, and their conceptualising and application of Sustainability Assessment is similar to the framing of the NIWA/Cawthron experimental concept, with its identification of impacts on receiving waters (environmental capital) of current stormwater management practices and the reflection of these impacts in ecological, social and financial values. In short, decision making aids such as Sustainability Assessment should incorporate explicit consideration of limits and targets as they relate to the management of capital stocks (see next section), and most likely need to recognise that some of these limits may already have been transgressed in particular locations or may be at risk of being transgressed.

Identifying objectives to be applied in the assessment - sustainability concept definition

The step change from earlier versions of assessment practice is the shift from relative assessments which focus on assessing whether or not proposals or policies result in outcomes which are generally in the desired direction ("direction to target" assessments) to absolute or comparative assessments which focus on assessing how far a particular proposal or policy will steer outcomes towards desired targets ("distance to target" assessments) (see, for example Pope et al., 2004).

For such assessment procedures, objectives have to be defined in a manner which renders them amenable to quantification or qualitative comparison.

With regard to practical methodologies for assessment and decision making, the central debate here is between those who adopt a 3-pillar[11] approach to defining objectives and associated criteria and indicators, and those who attempt to define sustainability objectives in a different and perhaps more 'integrated' way.

In this debate, Sadler (2008) represents the 3-pillar approach, including the concept of the *"stock of assets ... available to meet human needs"* (p.10) -

> *"In this concept, aggregate development is non-sustainable if the opportunities represented by net capital wealth are being depleted or eroded, but sustainable if they are being maintained or increasing on a per capita basis (while also reducing intra-generational inequity)."*

[11] In the international literature, '3-pillar' refers to the commonly referenced three pillars of sustainable development - economic, social and environmental well being. In New Zealand, it might be adapted to be called '4-pillar' reflecting the addition of cultural well being.

This over-riding sustainability objective then requires several subsequent steps which give more detailed expression to objectives, including -

- agreeing the 'level of sustainability' to be aimed for:
- specifying a set of more detailed objectives, each with their own outcome indicators;
- applying this set of objectives against either a triple top line (TTL, or *"normative values to aim for"*) or a triple bottom line (TBL, or *"safe minimums to stay within or warning signs to avoid"*), or a combination of the two - i.e. using *"a 'ceiling to floor' framework"* (p.12); and
- aiming overall for a net gain in sustainability outcomes.

In Sadler's approach, integration is achieved when economic, environmental and social factors *"are addressed simultaneously, and evaluated against a sustainability framework derived from international or national policy or strategies."* (p.11)

Gibson (2006) and Kemp et al (2005) represent the alternative approach to specifying objectives. Gibson's recommendations are made in the context of discussing what he sees as *"the basic sustainability requirements that should inform a transition to sustainability assessment"* (p.170); the inference being that other formulations (like that outlined above) have not made this transition.

His basic proposition is that sustainability objectives should be defined in terms of the observed *"requirements for improvement rather than the established categories of expertise"* (p.173). While acknowledging that the pillar-based approach has helped communicate the notion of inter-dependency, he points to the associated risk that this approach *"perpetuates fragmentation"* particularly fragmentation of decisions, actions and responsibilities by different agencies. His procedural suggestion is expressed as follows -

> *"Bottom-up sustainability assessments, driven by the expressed public concerns surrounding particular cases or initiatives, often <u>abandon the pillar categories and focus instead on problems and aspirations</u> that cross the social/economic/ecological boundaries. Public-issue identification and priority setting processes typically identify secure livelihoods, safety, health, vibrant and attractive communities, new opportunities and choice, and influence in decisions as key objectives. None of these is a purely social, economic or ecological matter."* (emphasis added)

Reflecting on professional experience and the literature, Gibson proposes a set of *"basic sustainability requirements"* (presented here in Appendix 3). Gibson's approach emphasises the following attributes about the list of basic sustainability requirements -

- they are not pillar-based, although they are linked to each of the pillars;
- they focus attention on *"what must be achieved, and what key actions are involved, to move consistently towards greater sustainability"*;
- they are not intended as a prescriptive formulation for everyone - it could be *"reconstructed, reordered and reworded in a host of different ways"*;
- they should be subject to continual review.

In Gibson's approach, integration is achieved by avoiding the traditional fragmentation into pillar-based analysis, and thereby avoiding the presumptions of responsibility for action that might flow from such fragmentation. In this regard, this approach may be seen as a stronger challenge to old habits of decision making.

The two approaches contrasted here have some attributes in common. Both promote the clear articulation of an over-riding sustainability objective, supported by a hierarchy of general and more specific objectives, some of which are expressed as criteria for assessment and decision making. Both approaches place great importance on the significance of trade-off rules (see next section) and the relationship between assessment and decision making (see further below).

To this point in time, the pillar-based approach has probably been trialled more, where as the approach articulated by Gibson (amongst others) is more an indication of a theoretical ideal inviting further experimentation and trialling.

The NIWA/Cawthron research, involving choice modelling, could be interpreted as moving beyond the pillar approach by focussing on “requirements for improvement” in terms that are recognisable to members of the public - underfoot conditions, water clarity and ecological health. It also encapsulates elements of the approach articulated by Sadler - the simultaneous consideration of environmental, social and economic factors, and may incorporate the notion of upper and lower thresholds (sustainability targets and limits), depending on how the low/medium/high ranges are specified.

Clarifying trade-off rules - sustainability concept application

A common theme in literature on sustainable development is the question of balance in the simultaneous pursuit of objectives (for example, Videira et al. (2010, p.447); Gibson (2006, p.172); Kemp (2005, p.21); Listowski et al (2009, p.84); Brinsmead (2005, p.151); Sadler (2008, p.18) and many more). Far less common is a discussion which provides any detailed guidance as to how the question of balance should be approached.

Both Sadler and Gibson are critical of the notion of compromise where one valued outcome may be achieved at the expense of another. When discussing overall outcomes - *“net determination of gains and losses”* - Sadler (2008, p.18) states the following view -

> *“the process of arriving at such determinations is always difficult and often controversial; trade-offs are a matter on which reasonable people holding different values and interests are likely to disagree. However, within a sustainability framework, the important point is that an explicit, systematic and good-faith attempt is made to strike the best possible balance”*

Gibson (2006a, p.172) takes what appears to be a more absolute stance -

> *“Sustainability is not about balancing, which presumes a focus on compromises and trade-offs. Instead the aim is multiple reinforcing gains. Trade-offs are acceptable only as a last resort when all the other options have been found to be worse.”*

However, he later (p.173) tempers this stance with some pragmatic acceptance -

> *“For sustainability, positive improvements are needed to meet all of the core requirements. Each is crucial and all are to be applied together. Significant and lasting improvements rely on linked, mutually sup-porting, positive steps on all fronts. There is no way around this. In practice, however, compromises and trade-offs will be unavoidable in most policy, program, plan and project decisions, if only because overall global conditions are now so very far from sustainability.”*

Videira et al. (2010, p.447) re-cast the assessment process as a shared problem-solving exercise rather than a project- or policy-approval exercise -

"To produce transformational outcomes, ISA processes focus on how the relationships underlying undesirable trends may be changed, and not directly on the impacts of a previously defined proposal, as occurs in other IA conceptions."

Nevertheless, it is difficult to avoid the concept of trade-offs, and the related questions of how should trade-offs be conceived in a sustainability context, and under what circumstances should they be considered.

Under the pillar-based approach articulated by Sadler, he addresses trade-off rules at two levels. First (p.11) are the rules which allow or disallow trade-offs between natural capital and economic capital under the three objectives for sustainability[12] - 'weak sustainability', 'moderate sustainability' and 'strong sustainability'. Second are the rules *"for trade-off and decision-making for sustainability assurance"* (pp.18-19) which address time and space dimensions of outcomes[13] as well as the relationship between 'net effect (gain/loss)' and decision criterion. For clarity, the latter is presented in Appendix 3.

Sadler then lays out *"a set of general and enabling guidelines"*, linking these to various sources in the sustainability assessment literature. The three guidelines are summarised below[14] -

"(1) At all stages of decision making, priority should be given to options and actions that do the most good, then to those that do no harm, and finally to those that accept some adverse effects.

(2) In principle, all other configurations of choice would be unacceptable within a sustainability framework, strictly defined.

(3) On some level, hard choices and trade-offs are an inevitable part of decision making. This task must be confronted rather than circumvented ... It should be undertaken in a transparent and accountable manner."

Perhaps unsurprisingly, Gibson's approach to trade-off rules for sustainability assurance is similar to Sadler's, even if it is laid out differently. He also places more emphasis on who is making the trade-off decisions, observing, as Sadler does, that even when there is apparent consensus on overall sustainability objectives *"different interests are likely to reach different conclusions about which of these compensations and net calculations may be justified."* (p.175).

Gibson argues for one essential general rule - *"trade-off decisions must not compromise the fundamental objective of net sustainability gain".* Beyond that, few rules will apply everywhere, implying the need to consider particular contexts. He suggests that other general rules to consider might include -

- no significant trade-offs or compromises permitted unless approved by all relevant stakeholders;

- no net additional burdens on the poor or already disadvantaged;

- no significant adverse effects of any kind can be justified by compensations of other kinds or in other places.

Gibson explains the need for process rules as well (p.175), observing that it is important to

[12] Full text is provided in Appendix 4.
[13] Outcomes across each pillar 'here and now' and 'elsewhere' and 'later'.
[14] Complete prescriptions are provided in Appendix 3.

ensure that the difficult choices are approached in an acceptable way. Such process rules include that -

- deliberations should be open to scrutiny and participation by interested parties;

- rationales be provided for proposed trade-offs - *"clearly identified, openly discussed and explicitly justified"*.

Without knowing more about the details and assumptions of the choice experiment scenarios envisaged by the NIWA/Cawthron research team, it is not clear what trade-off assumptions are embedded within the overall decision logic. However, the UPSW Progress presentation on 26 April 2010 did state that "Total Economic Value may be inferred from tradeoffs made in choice experiments" (slide 7). What can be said in terms of an injunction to good practice is that any trade-off assumptions and mechanisms that may be <u>implicit</u> within the choice experiments and associated modelling should be made <u>explicit</u> and discussed with the decision makers to ensure that they are aware.

Adaptive process design in Sustainability Assessment

Two aspects of 'adaptiveness' were encountered in the literature. The most common interpretation of 'adaptiveness' relates to iterative assessment processes which have evaluation and learning procedures built into them. (Kemp (2005, p.17, p.23); Gibson (2006a, p.173); Videira et al. (2010, p.453); Tuinstra et al. (2008, p.135); Brinsmead (2005, p.23); Sadler (2008, p.30)). The other interpretation of 'adaptiveness' refers to the need to customise assessment processes to the local context. (Gibson (2006a, p.172); Sadler (2008, p.25); Thevenot (2008, p.1); Deutsch and Metelka (2008, p.16))

For example, Gibson (p.172) notes -

> *"The notion and pursuit of sustainability are both universal and context-dependent. While a limited set of fundamental, broadly applicable requirements for progress towards sustainability may be identified, many key considerations will be location-specific, dependent on the particulars of local ecosystems, institutional capacities and public preferences."*

These two interpretations can of course be seen as complementary - where learning from the assessment process enables better customisation of the assessment process to the local context. Iterative, in-built learning processes will be discussed further in section 3.2.5 under Social Learning.

To the extent that the NIWA/Cawthron choice modelling experiment is being applied at three different types of locations - upper harbour, middle harbour, outer harbour - the contextualising/customising mode of adaptiveness may be relevant. For example, it might be asked whether the 'underfoot experience' is likely to be as relevant in the upper harbour and middle harbour locations as it is in the outer harbour location. Similarly, are users in these different locations likely to have the same set of relative preferences.

The relationship between assessment and decision making

The critical issue here for good practice in Sustainability Assessment is the relationship between the activities of assessment and decision making. Are the decision makers themselves involved directly in the assessment process and thereby intimately familiar with the mix of information and value judgements which have informed the results?

Boulanger (undated, p.3) refers to this as *"institutional integration"*. He uses this term to describe how Sustainability Assessment may be linked into policy making, and elaborates in terms of two possible meanings: a legal requirement for policy makers to carry out Sustainability Assessments (or have them carried out) or a procedural integration of Sustainability Assessment into the policy-making process. Unfortunately, on the latter, he does not go into detail.

In contrast, both Gibson (2006) and Sadler (2008) express unequivocal views that the direct involvement of decision makers in the assessment activities is a central element of integration in the context of Sustainability Assessment.

Gibson expressed clearly his interest in the design of an appropriate Sustainability Assessment regime, implying a degree of institutionalisation of the Sustainability Assessment processes. In line with this thinking, he stated that -

> *"The processes must put application of these sustainability-based criteria at the centre of decision-making, not as one advisory contribution among many."* (p.172)
>
> *"In conventional decision making, trade-offs between narrowly biophysical or ecological considerations and competing social and economic objectives may be made outside the assessment framework. In Sustainability Assessment, all the policy commitments and all the development objectives are considered together and the trade-offs are addressed directly"* (p.173)

There is a strong sense of direction towards collaborative governance arrangements - multi-stakeholder involvement with problem definition, objective-setting, options-assessment and decision making by the same group of people and represented interests. To separate assessment from decision making defeats the aim of integration "because trade-offs may often be necessary but should always be a last resort and should be subject to certain rules, otherwise they become the very instrument of defeat." (p.172).
Sadler (2008, p.11) similarly describes the linkage between assessment and decision making as a central aspect of integration stating that -

> *"An integrated approach calls for three types of integration: - **Substantive integration** of the three pillars and their different types of impacts within the assessment and decision-making processes*
> *- **Process integration** of the different stages, steps and methods of assessment*
> *- **Strategic or policy integration** of the inputs of assessment into decision-making process on a proposed action and/or into the larger policy, planning or project cycle"*

This procedural issue is clearly critical to several aspects of Sustainability Assessment, including setting objectives, specifying trade-off rules and determining priorities among criteria. Kemp (2005, pp.19-20) reinforces the same point from a policy governance perspective. While stating clearly that *"Policy integration is not the consolidation of policies to create a single integrated policy dealing with everything"*, he does see important responsibilities for governance bodies in coordinating the *"use of long-term sustainability strategies for sectoral domains"* and ensuring integration of the associated *"communication programmes, and national action plans with targets and ongoing programmes for assessment, feedback and revision and conflict resolution procedures."*

It is not clear from the descriptions received so far of the NIWA/Cawthron choice modelling experiment what relationship is intended between those whose values are expressed in the choice experiment forums and those who will subsequently make decisions using the resulting Adaptive Decision Support System.

3.2.5 Stakeholder engagement in integrated assessment processes

As noted in section 3.2.2, one of the strong themes associated with integrative practice in the 2006 review was an emerging recognition of the importance of stakeholder participation to various dimensions of integration. This review has found that the focus on stakeholder participation has attracted much attention since then, resulting in a more sophisticated discourse about rationales and approaches (see for example, Tuinstra et al. (2008,p.129); Videira et al. (2010, p.449); Kemp (2005, p.18); Lai et al.(2008, p.318)).

Two themes are evident. The first is a greater differentiation and specification of the rationales, roles and benefits of stakeholder involvement in assessment activities. The second is a focus on the phenomenon of 'social learning' that has been experienced in many instances. Indeed, Tuinstra et al., referencing numerous other sources from 1989 to 2005, write in terms of -

> *"a long line of developments that have seen, on the one hand, a deepening and widening of formal stakeholder engagement in policy processes involving complex problems and, on the other hand, increasing interest in improving the effectiveness of social learning within participatory assessment processes."*

However, scanning the literature also reveals that perspectives and insights on stakeholder involvements can be influenced considerably by disciplinary background or point of entry. Involvement in Sustainability Assessment does not always equate with extensive knowledge of stakeholder engagement processes, nor does an interest in deliberative democracy necessarily associate strongly only with sustainability issues.

Greater differentiation and specification of stakeholder roles and functions in integrated assessment

Before considering the literature on stakeholder engagement, it is necessary to explain that the term 'stakeholder' can be interpreted to have several meanings. The general concept of 'stakeholder' refers to parties who have an interest in a particular proposition (e.g. a proposal to manage stormwater in a particular way), or who may be affected by the implementation of such a proposition (i.e. the stormwater management proposal will impact on them in some way, perhaps creating new opportunities or requiring changes in their behaviour). Different interpretations distinguish between organisations or agencies which may have <u>official or even statutory responsibilities</u> affected by a proposal (e.g. in the case of a stormwater proposal, this might include a regional council, a runanga, a territorial authority's water works department, the Department of Conservation, a port authority) and other organisations or individuals who may <u>see themselves as potentially affected</u> (e.g. individual residents or recreational users of the coastal area, coastal 'guardians' groups, local water recreation groups, landowners associated with diffuse stormwater discharges such as farmers, orchardists, grape growers, those with businesses activities which may generate risks to coastal water quality, such as marine engineering and boat building, and so on). Both these interpretations are evident in New Zealand impact assessment reporting. Furthermore, in the New Zealand cultural context of treaty relationships, iwi and hapu do not consider themselves to be 'stakeholders' in a formal sense, although they would invariably consider themselves potentially affected parties in any coastal water management issue.

In the discussion of stakeholder engagement which follows, an inclusive interpretation of 'stakeholder' is advised[15].

In the 2006 Review, Brinsmead's discussion of the roles of stakeholders pointed to involvements beyond contextual integration for the sake of legitimacy. Nevertheless, Tuinstra et al. (2008, p.133) re-iterate that this is still an important factor, describing stakeholder input as essential for two main reasons: the inherently normative characteristic of sustainability policy decisions means that experts have no particular mandate for deciding the 'best path' to a sustainable future; and the need to build public acceptance of any changes required in order to move towards a more sustainable future.

Kemp (2005, p.18) supports the legitimacy argument and adds three others: stakeholder involvement helps reduce the risk of conflict, offers an additional source of ideas and information, and enables people and organisations to learn about environmental problems (discussed further below).

Videira et al. (2010, p.449) write in terms of a paradigm shift, describing *"six participation prototypes"* -

> - stakeholder input as a means of getting all problem-relevant knowledge and values incorporated into decision-making processes;
> - stakeholder representation of public preferences;
> - introducing deliberative processes as a method for rational competition of arguments;
> - independent citizen juries applying common sense to decide conflicting interests;
> - providing an opportunity for less privileged groups in society to be heard and to be empowered to become more politically active; and
> - enlightening the policy process by illustrating the diversity of factual claims, opinions and values.

Mathur et al. (2007, p.6) capture a similar range of factors -

> *"In addition to resolving conflicts, capturing multiple forms of knowledge, facilitating enhancement of social capital and increasing social learning, ... from a democratic perspective, inclusive decision making can be considered as an end in itself. More democratic decisions are generally considered to be more acceptable and increase a feeling of empowerment."*

This range of justifications points clearly to the potential scope for stakeholder involvement throughout every stage of an integrated assessment process, from issue framing and objective setting, through deliberation and prioritising to decision making. Indeed, Gibson (2006, p.172) takes an instrumental approach when he suggests that stakeholders are important for establishing sustainability objectives, decision criteria and trade-off rules for local contexts, when dealing with value-laden preferences. Videira et al. add that in the context of scientific uncertainty, *"experts are seen as no better equipped to decide on questions of values, interests or acceptable levels of risk than any other group of citizens."*.

Videira et al.(p.449) provide a succinct description of the trends which have seen a progressive shift from expert, managerial processes towards deliberative democracy. Scepticism over managers' ability to identify public interest properly and calls for early public involvement led to the concept of pluralism, where government is the arbiter of interests

[15] Mathur et al. (2007, p.7) provide a tabular summary (Table 1) of different objectives of Sustainability Assessment and the implications of a narrow definition of stakeholders for achieving each objective.

within the public rather than the source of objective decision making itself. Pluralism has evolved into structured processes for deliberation where *"dialogues between adversarial interests are promoted, but unlike pluralism those interactions are viewed less as a competitive negotiation than as a vehicle to identify the common good and subsequently act on shared goals."* As a result, stakeholder involvement is viewed as important not merely for the opportunity it affords them to influence decisions, but also because it contributes to strengthening civic capacity and social capital (Beierle and Cayford, 2002).

Practice and guidance

Although the concept of stakeholder engagement has evidently become more sophisticated over time, Videira et al. (2010, p.450) are not confident that experience of effective practice has kept pace, citing two recent review studies[16] which both found generally poor levels of engagement in practice. However, reflecting on their experience and the literature they venture to articulate some questions to guide good practice in the design and implementation of Integrated Sustainability Assessment (ISA). Breaking the assessment into phases - preparatory phase and implementation phase - they recommend -

Preparatory phase:

- a Steering Committee of 'institutional actors' and/or a broader forum of stakeholders should take responsibility for organising and executing the process;

- key diagnostic questions for this phase:
    - is the information to be used accessible to and interpretable by interested and affected parties?
    - where is decision-making authority and who would implement any agreements reached?
    - are there substantial disparities across participant groups in their power to influence the process?

Implementation phase:

- key diagnostic questions for this phase:
    - does the process aim at co-production of knowledge and revealing/appreciating different values and viewpoints regarding the policy undertakings at hand?
    - does the process aim to expose conflicts and resolve them through consensus approaches?
    - does the process envisage innovative solutions to problems, as well as the identification of win-win options and implementation partnerships? and
    - will the process be set to progress towards reaching consent on a common decision or set of actions?

Videira et al. note various options for recruitment of both Steering Committee members and other stakeholder representatives, citing more extensive work by Kallis et al, (2007) and van den Belt (2004). Furthermore, they point to several other aspects that may be critical to successful engagement - having a peer review group to assess the quality of the assessment procedures, and skilled facilitation.

[16] Lopes (2008) analysed 19 SEA Reports in six different countries. Thiel (2009) analysed the use of modelling tools in Sustainability Impact Assessment.

Kallis et al (2007) *"recommend assigning facilitation to an experienced facilitator with no ties to the organizers and preferably, an outsider to the system. One or more facilitators may be needed depending on the nature of the process, size of the group and technical expertise in face of the breadth of tools to deploy."*

Videira et al. defer to others (Vennix (1996) and van den Belt (2004) who have specified the competencies needed in order to facilitate and mediate participatory model building processes.
Interest in stakeholder engagement processes is high and is attracting considerable attention and effort at the present time. This review can only confirm the importance of the field of interest to the research programme anticipated by NIWA/Cawthron. How the research team has already engaged stakeholders or intends to address questions of stakeholder engagement is clearly important, for all the reasons outlined earlier in this section.

The seminar hosted by the Ministry for the Environment in June 2010 is evidence that numerous research teams in New Zealand are currently focussing attention on various methodologies for participatory research activities involving stakeholders in a variety of programmes. It is recommended strongly that the NIWA/Cawthron research team initiates discussions with other research teams to explore issues and options in this area of common interest.[17] The literature makes it clear that initiatives on stakeholder engagement cannot be subject to rigid guidelines but need to be adaptive to the task at hand. Such initiatives inevitably involve an element of experimentation, informed by the accumulating practical experience of facilitated participation processes (that exists within the research and professional community) and a commitment to evaluate effectiveness periodically.

Social learning

Some of the authors with experience of stakeholder engagement processes also tend to have an interest in the shared learning processes that can be supported in such circumstances. Other points of entry into the field of social learning are to be found where particular analytical and deliberative tools have been used as a means for structuring learning opportunities (for example, Proctor and Drechsler (2006); Hermans et al. (2007); Salgado (2009). In the context of this review, such tools include mediated modelling and multi-criteria assessment (discussed in section 3.3).

As a starting point in this discussion, the keywords provided by Tuinstra et al. in their 2008 journal article entitled "Learning and evaluation in Integrated Sustainability Assessment" are instructive of the potential scope of cross linkages between concepts: social learning; sustainability learning; 'learning-by-doing';conceptual learning; issue reframing; transformative outcomes; evaluation of learning.

More specifically, they explain the concept in the following terms -

> *"Social learning involves participants learning alongside each other about issues, but also about each other in respect of different perspectives on the issues and the basis of these different perspectives, so providing opportunity for mutual understanding and mutual influence. Social learning may enhance the capacities of those involved in a participatory assessment process (especially those who hold actual or potential agency in respect of problem causation or solution) to change the original institutional conditions and contexts, which 'frame' the issues and which potentially limit (unnecessarily) the possibilities to find solutions."* (p.131).

[17] Of particular interest might be dialogue and exchange with Dr Marjan van den Belt at Massey University who has considerable experience in the field of mediated modelling for adaptive management.

Mathur et al. (2007,p.6) emphasise the validity of different types of knowledge -

> *"The concept of social learning, as compared to technical learning, derives from the argument that different kinds of stakeholders including the layman have an equally valid knowledge. And it has been argued that the pursuit of sustainability requires enabling such processes which facilitate social learning."*

Gibson (2006, p.172) refers to 'experiential learning as being fundamental to the (non-pillar) reframing which he argues for, citing what has now come to be accepted by most as a truism - "experiential learning that there could be no species preservation without habitat preservation and no habitat preservation without local livelihood security".

Of particular interest here are the references made by a number of authors to the use of Multi-Criteria Assessment/Evaluation (or similar descriptors) as tools to facilitate social learning. For example, Proctor and Drechsler (2006, pp.172-3) suggest -

> "*MCE has the advantage of being able to provide a framework for complex decision making problems that allows the problem to be broken down into workable units and to be structured in such a way that the complexities of the problem can be unravelled. This is done essentially through the process of identifying options, criteria, and preferences. Applying an MCE in a heuristic way enables the MCE to aid in the learning process concerning complex issues."*

Similarly, Hermans et al. (2007, p543), referring to the use of a Multi-Criteria Decision Aid to assist a multi-stakeholder group to evaluate proposed river management options, pointed out that *"the process was used primarily as a social learning tool."* They added that the structure of the process would be similar whether the outcomes were being used primarily to facilitate discussion and group learning or to achieve consensus.

Weaver et al. (2006) highlight important deliberative characteristics of social learning in the often highly contested field of Sustainability Assessment. Situations of conflicting values and competing world views need to be used to stimulate the joint search for (sometimes) innovative solutions rather than outcomes where some gain at others' expense and conflicts are reinforced. They express the role of social learning as follows -

> *"is an open-ended process in which individuals and groups continuously learn how to frame and reframe the issues at stake in a more socio-ecologically robust way, how to integrate the potential for innovation and creativity derived from social conflicts, how to deal with competing world views in a constructive and cooperative way, and how to create social collective capacities to deal with common problems."*

Few papers were retrieved that delved into the epistemology of social learning per se. Tuinstra et al. (2008) provide some guidance on conditions affecting social learning and on approaches to evaluating the extent of social learning, which are of potential relevance in the research context of this review. Drawing on the evaluation of three integrated assessment processes *"where the attempt has been made to institutionalise learning procedures into the process design"* (p.132), they report that participants identified the following key conditions that affect social learning -

- the motivation and skills of leaders and facilitators;
- clarity about the role and purpose of stakeholder involvement;
- the internal structure (of the group) and the room given to democratic debate about regulatory institutions;

- the structural capacities[18] for interaction between social networks; and
- the influence exerted by the existing culture on the framing and definition of the issues at stake.

Tuinstra et al. (2008, p.134) highlight the case of the Georgia Basin Futures Project, reported in Robinson et al. (2006) -

> *"based on the assumption that the choice about what sustainability means in concrete terms for a given context and group emerges from a social learning process, where expert knowledge is combined with stakeholders' attitudes, preferences and beliefs."*

The lessons learned about project design and documented in Robinson et al (2006) appear germaine to the present discussion, and are set out below -

> *"• as much time and thought needs to be given to the design and management of the modelling workshops as to the model development process itself;*
> *• key to the workshop processes is that scenarios are created by the users and not by the experts;*
> *• participants need a lot of time to learn, understand and use tools:*
> *• regional sustainability is an extremely complex issue and a model can at least provide useful shortcuts and heuristics that help people understand the nature of interactions between the subsystems it seeks to represent;*
> *• communicating complex information in formats that are simple and easy to understand is a challenge;*
> *• interactivity and visualisation can be very powerful ways of engaging users;*
> *• different stakeholder and user groups may have very different expectations of tools and different views of the goals of workshops;*
> *• the analysis of the consequences of particular technological and behavioural choices (the realm of scenario analysis) has to be separated from the discussion of the desirability of those outcomes and the means that may be required to realise them;*
> *• participants found it difficult to distinguish between desirable and feasible policy interventions:*
> *• there can be a tension between engaging stakeholders in thinking through alternative futures and the changes that might be required to get to those futures, and the (missing) mandate of the scientific team for making any of those changes happen:*
> *• there can be an unavoidable tension between, on the one hand, the desire to engage partners and users actively and to contribute to a process of social mobilisation around sustainability issues and, on the other hand, the desire to do academically credible research."*

To test if and where the constructive transformations expected as an outcome of social learning occur, Tuinstra et al. (2008, p.143) suggest that -

> *"Evidence of learning may come in the form of 'stated' changes in understandings, beliefs, goals, or strategies (such as may be found through interviews with stakeholders and scientists) or 'revealed' through indirect evidence (such as by reframing of the problem and solution, changes in the scope of the assessment, shifts in discourse, new questions, and new experiments.)."*

They elaborate some principles for evaluating learning by proposing three broad areas of learning to focus on -

- <u>social and sustainability learning within the issue domain</u>: *"learning about the content and substance of sustainability and the issues and undertakings being*

[18]Taken to mean the balance of influence and power relations within the group.

assessed and each other's perspectives on these" and noting that this area of learning mainly concerns *"stakeholders and actors and has potential to change their behaviours"*

- policy- and strategy-relevant learning concerning the issue domain: *"the relevance of findings for development policy, transition management, science policy, assessment policy, etc."* and noting that this area of learning mainly concerns *"those with agency in implementing (policy) solutions and has potential to change policy frames, policies, institutions, business strategies and agendas"*
- methodological learning about Integrated Sustainability Assessment: *" the new ideas and concepts it introduces, the demands it makes, the limitations it exposes in the existing tool kit and in conventional ways of working, the implications it holds for the science agenda, etc."* and noting that this area of learning mainly concerns *"project managers and scientists and has the potential to lead to institutional change within science and changes in science agendas".*

In light of the foregoing discussion, achieving social learning outcomes amongst participating stakeholders and the NIWA/Cawthron research team will be an important test of the effectiveness of any resulting ADSS.

3.3 Requirements for developing an integrated sustainability indicator system (specifically an ADSS based on a Multi-Criteria Assessment tool)

3.3.1 Overview

This review has accessed a number of Multi-Criteria Assessment (MCA) cases being applied in the context of Sustainability Assessment and related decision making. Most of these are cases of trialling a general methodology to a particular sustainability challenge. Within this selection of literature, a number of authors referred explicitly to the use of a MCA methodology as a means for structuring stakeholder engagement - see for example Proctor and Drechsler (2006), Hermans et al. (2007); Salgado et al. (2009), Soma (2010), Deutsch and Metelka (2008), Lai et al. (2008).

In light of the findings on good practice in Sustainability Assessment, these cases have been analysed in terms of -

- MCA structural details (objective(s); alternatives; principles for criteria selection; categories of criteria; weighting of criteria; assessment procedure; aggregation procedure; sensitivity analysis);

- stakeholder engagement process (stakeholder type; stakeholder selection; relationship of stakeholders to ultimate decision makers; use of expert facilitator; steps in stakeholder group process; governance arrangements); and

- expressed[19] strengths, weaknesses and issues of each MCA application.

Analytical summaries are presented in tabular form in Appendix 4. In section 3.3.2, the expressed strengths and weaknesses are reviewed to see if they provide additional insights to those already gleaned from the broader Sustainability Assessment literature and reported in section 3.2.

[19] i.e. expressed by the authors.

The DayWater Project has been a major international research and development initiative[20] attempting to develop an Adaptive Decision Support System (ADSS) for urban stormwater source control, an objective similar to that of the NIWA/Cawthron research team. Of the cases reviewed, the DayWater Project is unique in attempting to produce a computer-mediated ADSS tool for remote access by users. All the other cases involve the use of MCA methodologies in interactive group processes, although some of these involved use of computer-based tools mediated by the researchers.

Section 3.3.3 will review the experience of the DayWater Project from the reported experiences of the MCA tool developer and a prospective end user to see what additional guidance can be gleaned for the NIWA/Cawthron ADSS development work.

3.3.2 Expressed strengths, weaknesses and process issues encountered in MCA applications

Strengths:

The strengths of MCA processes from an expert perspective are associated with their contribution to clarifying logic and the fact that they allow values to be incorporated into assessments in an explicit manner.

For clarifying logic, it is claimed that -
- MCA processes allow a complex problem to be broken down into workable units, where options, criteria and preferences are made explicit;
- simpler MCA methods can encourage transparency of logic (valued by stakeholders) and provide a useful structure for communicating decisions;
- the MCA framework assisted in structuring preparations and deliberation processes; in making information transparent (alternatives, criteria, impact scores, weights) and pointed to a clear structure for each participatory process.

For getting values incorporated, it is claimed that -
- MCA processes result in a good combination of agreed facts and social values;
- citizens were involved to make comparative values judgements in a long-term perspective;
- citizen juries can be used to aggregate multiple individual preference weights through deliberation to achieve consensus.

The strengths of MCA processes from stakeholder/participant perspective are associated with the perceived benefits of structured and open deliberations, the perceived legitimacy of their participation, and the opportunity to learn and understand complex issues better.

The benefits of structured and open deliberations experienced are described as -
- giving citizens increased understanding of different points of view;
- enabling the group to learn and ‘move forward’;
- encouraging participants to focus more on the preferences and weightings for the criteria than on the final outcome; and
- while framing of citizen deliberations favoured social values, the format did not preclude voicing of individual interests - but did tend to de-legitimise them.

The perceived legitimacy of participation is reflected by -

[20] Involving 10 scientific partners and 14 end-user organisations from seven European countries over a period from 2002 to 2008.

- the involvement of interest groups (clarifying different interests) and citizens groups (comparison of interests) being seen as complementary;
- the method of interest group involvement focussing on explanations of their perspectives rather than lobbying for specific solutions.

Enhanced learning is provided for -
- once the MCA tool is familiar to users, the ease and speed of use allows a degree of ‘experimenting’ with weights to test concerns and options, or when applied in a heuristic (iterative with feedback) way, the MCA tool aids jury learning about complex issues; and
- the use of maps and visual illustrations of alternatives to aid transparency.

Weaknesses:

The weaknesses of MCA processes from an expert perspective are associated with data problems, methodological challenges, process facilitation difficulties and resourcing needs for the assessment.

Data problems may be associated with -
- difficulty in finding quantitative and/or accurate information on many of the criteria identified for assessment;
- difficulties with identifying or agreeing scale descriptors for impacts.

Methodological challenges arise -
- when it is not clear whether the method has dealt appropriately with trade-offs or compromises involving compensation of one factor loss by another factor gain (see recommendation 1 in section 4.1)
- over the assumption that preference for different criteria are assumed to be independent of each other;
- double-counting problem when chosen criteria are either redundant or non-exhaustive.

Interactions between analyst and decision makers can become difficult if the analyst is taking on dual roles of science expert and process facilitator (see recommendation (vi) in section 4.1).

The issue of resources needed to undertake an assessment process (or in this case a development process) is beyond the scope of this review. However the resourcing implications which arise if many assessment criteria and/or stakeholders are involved must be taken into account. Stakeholder engagement processes need to be ‘fit for purpose’ within the constraints of the resourcing available.

The weaknesses reported from a stakeholder/participant perspective appear mainly to be associated with excess complexity and an absence of effective social learning. They include -

- the need to balance complexity/simplicity with cognitive capacity; complex MCA methods can be perceived by non-experts as “black box” approaches; too many objectives/criteria can overload individual’s thinking and analysis;
- citizens can become overwhelmed by expert contributions in some situations - distracting them from their long-term focus;
- the differences between scientific knowledge and practical knowledge and the different ways of thinking about a real/abstract situation or problem;

- having experts/interest group representatives select criteria risks missing some criteria considered important by individual citizens;
- where citizens juries have not been involved in the task of structuring the decision making.

3.3.3 Lessons from DayWater experience

Whilst the NIWA/Cawthron concept is not identical to any other approach encountered in this review, the intention to develop an ADSS, means that lessons from the DayWater experience are likely to be of interest.

The documented compendium of papers describing many aspects of the DayWater Project over its progress from 2002 till 2008 is available to the NIWA/Cawthron research team. It's details will not be reproduced here.

At the level of programme design it is particularly useful to have had access to papers written from several perspectives, notably the scientific/technical perspective and end-user perspective. The scientific/technical perspective addresses, amongst other aspects, the design of the overall ADSS architecture (Deutsch and Metelka, 2008) and the specific MCA tool (Ellis et al., 2008), in this case referred to as the DayWater Multi-Criteria Comparator (MCC). The end-user perspective (Beyeler, 2008) describes an end user's account of their involvement in and experience of the prototype development.

Some contextual detail should be made clear in order to put any other observations into context. The DayWater Project was an immense and ambitious undertaking. The Project involved a consortium of 10 scientific partners, from the polar circle (Sweden) to the Mediterranean Sea (Greece). Thevenot (2008, p.2) describes the consortium as *"dominated by engineers and urban hydrologists"* but also with *"experts in socio-economics and information technology"*. There have also been 14 core end-users, with the intention to extend to a much wider group of end users for testing components of the ADSS[21] Core end users are predominantly state, city or county water agencies, indicating a narrow conception of stakeholders. For geographic and linguistic reasons; local DayWater partners have been responsible for interacting with their Core end users. Thevenot (p.5) reports that end-user involvement was one of the most difficult tasks, constrained by time demands, language barriers, and the level of mutual understanding achieved. Collecting and collating end-user needs in order to agree the functionality of the ADSS was much more demanding than expected, taking 24 months instead of the expected 12 months. Clearly, project coordination was problematic, facing difficulties of multi-disciplinary and multi-national dimensions -

> *"even if scientific partners are working in the same field, there is a common language to be found for the project. ....Each partner developed their components mostly independent of the other components and their mutual requirements. .... The analysis of the decision making procedure in urban stormwater management projects is a good example of this kind of intercultural and interdisciplinary misunderstanding. Consulting engineers generally conceive the decision making procedure in a different way than researchers and south European countries do not consider decision making procedures as north Europeans do".*

Thevenot's description (pp.5-7) does seem to imply a top-down, directive development process which may have acted as something of a straightjacket to innovation. He notes, for example that *"a strong motivation is needed to convince the numerous stakeholders involved in any urban stormwater management project of the effectiveness and sustainability of*

[21] Although the extended end user groups have been involved in a series of 8 regional conferences, it appears that they have yet to become involved in actual testing of ADSS components.

relevant 'Best Management Practices'". On the face of it, this appears as something of an admission of defeat when the aim was to develop an adaptive DSS. It seems debatable whether the involvement of such a wide variety of organisations in so many different countries and climatic settings actually works in favour of developing adaptive capacity - through consensus and standardisation at each stage. The question might well be asked whether or not it might have been preferable to cultivate diversity of approaches, with periodic group reflection and sharing of experiences amongst different groups of innovators.

These observations should not be taken to imply that the DayWater Project has not made substantial progress and created a DSS with numerous useful functionalities (library function, management function, analysis function and communication function). There is undoubtedly a web-based MCC tool available to its end users. However, it has many of the hallmarks of a smart "black box" when encountered by an uninitiated prospective user.

The critical question, in the context of this review, is whether the DayWater experience offers any useful guidance to the NIWA/Cawthron research team. In terms of technical, conceptual and analytical detail within the various ADSS model components, the answer may well be yes, although that is not for this review to determine. But in terms of guidance on the concept development process, the answer is far from clear.

For the record, the DayWater MCC tool assists a user to select from a range of 15 structural BMP options to address stormwater problems on a site with certain defined physical characteristics. Sixteen indicators across six categories of criteria[22] are available to guide users' assessments, making the dimensionality of the DayWater tool considerably more complex than the tool envisaged in the NIWA/Cawthron research. Users are able to choose how they weight the various criteria, with the sum of all weights equalling 100%. Performance scores for each criterion can either be across three bands (low, medium, high) as in the NIWA/Cawthron concept, or on a scale of 0-10. The interactive, computer-based tool allows users to engage in some individual experimentation or to generate several sets of results as the basis for discussion and group learning.

Beyeler (2008) providers an end-user perspective, representing one of the 14 Core end users. Her paper makes it clear that she is not an uninformed member of the public, as indicated by her analysis of the needs for improvement in the management of urban stormwater in the face of unsustainable trends (p.166). Beyeler points out that the launch of the DayWater Project created high expectations, including (p.168) -

> *"- a tool easy to use by urban stakeholders (engineers, technicians, elected officials, planners, developers, etc.)*
> *- a tool offering methodological advice, case studies, scientific, technical and financial information;*
> *- a sharing of experiences, both good and bad;*
> *- an incentive to think globally;*
> *- a dynamic database;*
> *- an adaptive approach allowing their request to be site specific."*

Beyeler then describes several experiences of interactions with the Hydropolis wesbite which indicate how the DayWater tool has either been made more user friendly through end-user feedback, or might be further improved. It appears from her user perspective that the approach to developing the ADSS has been technocratically driven by 'experts' in water management and information technology, but that the development process has not taken

[22] Technical, environmental, Operational & Maintenance, Social & urban community benefits, Economic costs, legal & urban planning.

non-expert end users along with it and has therefore been experienced as somewhat inaccessible at times. At other times, she and other end users have also been excited by their involvement. She argues for more direct involvement of end users in the testing and trialling of the ADSS, but their involvement has to start at the beginning - articulating how end users experience problems and ask questions, and letting the ADSS developers hear and respond to these.

Notwithstanding the apparent differences in research setting and scale between the NIWA/Cawthron research project and the DayWater Project, key lessons to be taken from the DayWater Project include -

- managing expectations, particularly in deciding which user needs to address and how users express their needs;

- working hard to establish a common language between scientists and non-scientists;
- maintaining continual dialogue and feedback between ADSS developers and users; and

- making sure that the research team does not function in a fragmented manner.

4 REVIEW CONCLUSIONS AND RECOMMENDATIONS

The following conclusions and recommendations are made to provide guidance to the development process of Cawthron's work with NIWA in the research programme on Urban Planning and Sustainable Water.

4.1 Part 1: Conclusions and recommendations from review of Sustainability Assessment literature

<u>Conclusions</u>

(i) The perspective on sustainability expressed in the NIWA/Cawthron research concept (a focus on the current status of receiving water bodies for urban stormwater and identified needs for improvement) aligns with the perspective expressed in recent applications of Sustainability Assessment.

(ii) The NIWA/Cawthron choice-modelling methodology appears to incorporate aspects of both predominant approaches to expressing sustainability objectives. The Total Economic Value concept in the modelling involves simultaneous consideration of social, ecological and economic dimensions while the three "environmental quality" attributes in the choice experiments reflect elements of value propositions that are not strictly pillar-based but are related to people's practical experiences of the coastline.

(iii) It is not clear what trade-off assumptions are embedded within the overall decision logic of NIWA/Cawthron's choice-modelling methodology.

(iv) At this stage of research concept development, there is no reason to expect that the intended ADSS will not be capable of both elements of adaptiveness - characterised by (i) iterative assessment processes which have evaluation and learning procedures built into them to enable adjustments in decisions over time, and (ii) the capacity to customise assessment processes to the local context. The latter is already provided for in the experimental design involving three separate locations/coastal water body sites.

(v) It is not clear from the descriptions received so far of the NIWA/Cawthron choice modelling experiment what relationship is intended between those whose values are expressed in the choice experiment forums and those who will subsequently make decisions using the resulting Adaptive Decision Support System (ADSS).

(vi) Contemporary practice in Sustainability Assessment points to the potential scope for stakeholder involvement throughout every stage of an integrated assessment process, from issue framing and objective setting, through deliberation and prioritising to decision making. In the NIWA/Cawthron research, these stages would include proof of concept through to ADSS development and trials. It is not known what provisions have been made for stakeholder engagement in the NIWA/Cawthron research programme, or to what extent this has occurred already.

(vii) There are as many as half a dozen research programmes in New Zealand currently exploring methods for incorporating competing values into decision making in water management, and therefore also likely to be considering various stakeholder engagement activities as part of their research. Some New Zealand researchers and consultants already have extensive practical experience in such matters.

(viii) The literature emphasises the importance of social learning to the contemporary practice of Sustainability Assessment. The literature also reveals that certain research tools, such as MCA and mediated modelling, can be helpful in structuring social learning processes. Achieving social learning outcomes amongst participating stakeholders and the NIWA/Cawthron research team will be an important test of the effectiveness of any resulting ADSS.

Recommendations

(i) Attention to trade-offs: good practice will require that any trade-off assumptions and mechanisms that may be implicit within the choice experiments and associated modelling should be made explicit and discussed with the decision makers to ensure that they are aware of the nature of the trade-offs associated with any choice/decision (related to Conclusion (i)).

(ii) Dialogue throughout the development process: building in both elements of adaptiveness to the ADSS will require close attention to stakeholder/researcher dialogue. In particular, building in evaluation and learning procedures to the methodology to enable adjustments in user decisions over time will require such dialogue to take place (be managed) at various stages of ADSS development, not just at the beginning and end of the development sequence (related to Conclusion (iv)).

(iii) Stakeholder selection: a selection of intended future users/decision makers should be included amongst the stakeholders involved directly and throughout the development of the ADSS. It should not however be assumed that they are the only stakeholders or indeed the most important stakeholders. (related to Conclusion (v)).

(iv) Timing of stakeholder engagement: if stakeholder engagement has not already been initiated, this should happen before the end of the ‘proof of concept’ stage with the expectation of periodic involvement throughout each stage of ADSS development. (related to Conclusion (vi)).

(v) Collaboration with other science programmes: NIWA/Cawthron should consider initiating a science forum devoted to stakeholder engagement issues, with a view to providing a regular forum for exchange of ideas and experience amongst scientists with these common interests. (related to Conclusion (vii)).

(vi) Expert facilitation: in order to maximise social learning during the research programme, the NIWA/Cawthron research team should consider the use of expert facilitators to plan and facilitate periodic group work sessions involving stakeholders. (related to Conclusions (vi, vii, viii)).

(vii) Testing for social learning outcomes: as a means of monitoring progress toward an effective ADSS, the NIWA/Cawthron team should consider monitoring social learning outcomes amongst participating scientists and stakeholders at intervals during the development process. (related to conclusion (viii)).

4.2 Part 2: Conclusions and recommendations from the review of selected MCA applications in Sustainability Assessment

<u>Conclusions</u>

(i) The use of just three attributes in the choice experiments will probably be an advantage in terms of striking the balance between complexity and simplicity. Increasing the number of choice attributes will increase user perception of complexity and time and resource costs of developing the tool.

(ii) The representativeness of stakeholders is important to a sense of legitimacy in deliberative processes (e.g. choice experiment forums, researcher/stakeholder deliberations).

(iii) It is important to avoid a situation where stakeholders perceive the MCA tool as a 'black box', because they have not been taken along with the process of its development. This reinforces recommendations (ii) and (vi) in section 4.1.

(iv) The data problems (data gaps; agreeing scale factors) identified in some MCA applications are addressed in the NIWA/Cawthron approach to attribute selection by requiring corresponding science measures that can be related to low, medium and high levels.

(v) In the use of ADSS, different functionalities may be of more or less interest to different users. This suggests the need to cast the net broadly at the stage when end-user needs are being defined and refined. Conversely, rather than try to make the ADSS tool "all things to all people", it may be preferable to restrict its functionalities in the first instance and concentrate on developing a useful tool which is better understood by a variety of interests. In short, managing user expectations as well as user education is likely to be critical to success.

(vi) It is important that the NIWA/Cawthron research team, working on different aspects of the ADSS development, do not operate in a fragmented manner.

Recommendations

(i) Composition of stakeholder interests: Consideration of the appropriate composition of stakeholder interests is important in order for the resulting ADSS to be viewed as a legitimate decision-making aid, and in order to meet the functionality needs of different users. (related to Conclusion (ii) and (v))

(ii) Expert facilitation: Use of independent facilitators in working sessions involving stakeholders and scientists will free up the scientists to focus on their roles as providers of expert knowledge. (see also recommendation (vii) in section 4.1)

REVIEW BIBLIOGRAPHY

Baines, J.T. and Morgan, B. 2006. Review of Integrated Impact Assessment methods and tools. Prepared for the Ecologic Foundation, with funding from the New Zealand Foundation for Research, Science and Technology. April 2006. 68p.

Beierle T. and Cayford J. 2002. Democracy in Practice. Public Participation in Environmental Decisions. Resources for the Future Press: Washington DC.

Beyeler, C. 2008. DayWater ADSS: an incentive for action but a source of questions - an end-user account of the prototype development. Chapter 16 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management.IWA Publishing, UK.

Booth, K, England, A, Rankin, D, Unwin, M, Charles, G, England, K, Riley, K and Ritchie, D. 2010. Whitewater Kayaking in the West Coast Region: Application of the River Values Assessment System (RIVAS). Report prepared as part of a project entitled 'Developing a significance classification framework for water body uses and values', funded by the Foundation for Research Science and Technology, Envirolink Grant 612-TSDC41. 14p.

Boulanger, P-M. undated draft. Integration in Sustainability Impact Assessment: Meanings, Patterns and Tools. Working paper as part of the research project Methodology and Feasibility of Sustainability Impact Assessment. Case: Federal Policy Making process , funded by the Belgian Science Policy under grant CP/F5/453.

Brinsmead, T. 2005. Integrated Sustainability Assessment: Identifying Methodological Options. Report prepared for the Joint Academies Committee on Sustainability. http://www.naf-forum.org.au/papers/Methodology-Brinsmead.pdf

Buselich K 2002. An outline of current thinking on sustainability assessment: A background paper prepared for the Western Australian State Sustainability Strategy. Institute for Sustainability and Technology Policy, Murdoch University, Western Australia. August 2002

Dalal-Clayton, B. and Sadler, B. (In Press) Sustainability Appraisal: A Review of International Experience and Practice. First draft of work in progress (January 2005), funded by the Royal Norwegian Ministry of Foreign Affairs.

Deutsch, J.C. and Metelka, T. 2008. Adaptive DSS - a new kind of DSS. Chapter 2 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management.IWA Publishing, UK.

Deutsch, J.C. and Metelka, T. 2008. DayWater ADSS guided tour: a feature for the Decision Support System adaptability to its user. Chapter 4 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management.IWA Publishing, UK.

Dewulf, A., François, G., Pahl-Wostl, C. and Taillieu, T. 2007. A framing approach to cross-disciplinary research collaboration: experiences from a large-scale research project on adaptive water management, in *Ecology and Society*, Vol. 12, No. 2, p.14,

Ellis, J.B., Revitt, D.M. and Scholes, L. 2008. The DayWater Multi-Criteria Comparator. Chapter 9 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management.IWA Publishing, UK.

Gibson, RB. 2006. Sustainability assessment: basic components of a practical approach. In *Impact Assessment and Project Appraisal*, vol.24, no.3, pp.170-183.

Gibson, RB. 2006. Beyond the Pillars: Sustainability Assessment as a framework for Effective Integration of social, Economic and Ecological considerations in significant decision-making. In *Journal of Environmental Assessment Policy and Management*, vol.8, no.3, pp.259-80.

Hermans, C., Erickson, J., Noordewier, T., Sheldon, A. And Kline, M. 2007, “Collaborative environmental planning in river management: An application of multicriteria decision analysis in the White River Watershed in Vermont” in *Journal of Environmental Management*, vol. 84, issue 4, pp. 534-546

Hughey, K, Booth, K, Deans, N and Baker, M. 2009. River Values Assessment System (RIVAS) - The Method. Report prepared as part of a project entitled ‘Developing a significance classification framework for water body uses and values’, funded by the Foundation for Research Science and Technology, Envirolink Grant 612-TSDC41. 14p.

Kemp, R. Parto, S. and Gibson, R.B. 2005. Governance for sustainable development: moving from theory to practice. In *International Journal of Sustainable Development*, Vol 8, Nos1/2. pp.12-30.

Lai, E., Lundie, S. and Ashbolt, N.J. 2008. Review of multi-criteria decision aid for integrated sustainability assessment of urban water systems. In *Urban Water Journal*, Vol.5 No.4, December 2008, pp.315-327

Listowski, A., Ngo, H.H., Guo, W.S., Vigneswaran, S. and Palmer, C.G. 2009. Concepts towards a novel integrated assessment methodology of urban water re-use. Paper presented at Challenges in Environmental Science & Engineering, 14-17 July, 2009, Townsville, Queensland, Australia.

Mathur, V.N., Price, D.F., Austin, S and Moobela, C. 2007. Defining, identifying and mapping stakeholders in the assessment of urban sustainability. International Conference on Whole Life Urban Sustainability and its Assessment. Horner, M., Hardcastle, C., Price, A and Bebbington, J. (Eds) Glasgow.

Ministry for the Environment, 2010. Weighing up competing values in decision making on freshwater quality and flows. Notes provided by the Ministry following the workshop held in Wellington 11 June.

Pope, J., Annandale, D. and Morrison-Saunders, A. 2004. Conceptualising sustainability assessment. In *Environmental Impact Assessment Review*, vol.24, no.6, pp.595-616.

Proctor W, and Drechsler M 2006. Deliberative multicriteria evaluation. In *Environment and Planning C: Government and Policy* **24**: 169-190.
Robinson, J., Carmichael, J., van Wynsberghe, R., Tansey, J., Journeay, M. and Rogersk, L. 2006. Sustainability as a problem of design: interactive science in the Georgia Basin, in *The Integrated Assessment Journal Bridging Sciences and Policy*, Vol. 6, No. 4,pp.165–192.

Russell, S and Ward, M. 2010. Sustainability Appraisal: Application to the Canterbury Water Management Strategy. Landcare Research Contract Report LC0910/052.

Sadler, B, Ward, M and Frame, R. 2008. A Framework for Sustainability Appraisal in New Zealand - Thinkpiece. Landcare Research Contract Report LC0708/090.

Salgado, P. et al. 2009. Participative multi-criteria analysis for the evaluation of water governance alternatives. A case in the Costa del Sol (Malaga), in *Ecological Economics*, vol. 68, issue 4, pp. 990-1005

Soma, K. 2010. Framing participation with multicriterion evaluations to support the management of complex environmental issues, in *Environmental Policy and Governance*, vol. 20, issue 2, pp. 89-106

Taylor, A.C. and Fletcher, T.D. 2005. 'Triple-bottom-line' assessment of urban stormwater projects. In proceedings of the 10th International Conference on Urban Drainage, Copenhagan/Denmark, 21-26 August 2005.

Thevenot, D.R. et al. 2008. DayWater: the challenge to develop an Adaptive Decision Support System for urban stormwater source control. Chapter 1 in Thevenot, D.R. (Ed.) 2008. DayWater: an Adaptive Decision Support System for Urban Stormwater Management.IWA Publishing, UK.

Toth FL 2003. State of the Art and Future Challenges for Integrated Environmental Assessment. *Integrated Assessment* Vol. 4, No. 4, pp. 250–264.

Tuinstra, W., Jäger, J. and Weaver, P.M. 2008. Learning and evaluation in Integrated Sustainability Assessment, *Int. J. Innovation and Sustainable Development*, Vol. 3, Nos.1/2, pp.128–152.

van den Belt, M. (in preparation). Multi-scale Integrated Modelling for Sustainable Adaptive Systems.

Videira, N., Antunes, P., Santos, R. and Lopes, R. 2010. A Participatory Modelling Approach to Support Integrated Sustainability Assessment Processes. In *Systems Research and Behavioural Science* Vol.27, 446-460.

Weaver, P.M. and Rotmans, J. 2006. Integrated sustainability assessment: What, why and how, in *International Journal of Innovation and Sustainable Development*, Vol. 1, No. 4, pp.284–303.

APPENDICES

Appendix 1: Interpretation of the brief and search parameters

Appendix 2: Core generic criteria for sustainability assessments - from Gibson (2006)

Appendix 3: Guidance on trade-off rules - from Sadler (2008)

Appendix 4: Analysis of MCA cases reviewed.

Appendix 5: Literature identified through internet and database searching

Appendix 1: Interpretation of the brief and search parameters

There are two components to the literature search -

1) To update the sustainability/integrated assessment review conducted for the Ecologic Foundation in 2005. The review should focus on post 2005 literature that contains pragmatic approaches and implementation details rather than polemics.

2) To search for literature that discusses approaches to implement an integrated sustainability indicator system for urban development scenario assessment in the context of receiving water bodies for urban storm water. The client has noted that “the integration device for the indicator system will be either a multi-criteria method or sophistications around the OECD (2005) composite index method, unless we discover an alternative approach”. A decision was made by the research team to focus on post 2000 literature. Further clarification by the client included:

- not to look at technical challenges of combining indicators from the socio-economic and biophysical domains;

- interested in adaptive governance requirements in social-ecological systems, indicators that assess the pre-requisites for transition to successful adaptive governance, indicators of resilience in the social-ecological system;

- ignore reductionist approaches such as OECD which focuses on generating one number;

- focus on governance and consultation aspects of MCA (multi-criteria approaches) rather than the nuts and bolts of how to calculate the overall decision (most MCA literature based on expert workshops where indicators are scored and weights applied, are there other approaches and how do they do it?);

- interested in sustainability assessment context (nuts and bolts/technical to be embedded into it).

Part 1: Update of the integrated/sustainability assessment review conducted for the Ecologic Foundation in 2005

Internet search engines/databases used:

- Google Scholar
- Directory of Open Access Journals
- Scirus
- Google

University databases used:

- Web of Science
- ProQuest
- ScienceDirect
- Informaworld
- Wiley Interscience

Keywords/phrases used:

- Sustainability assessment
- Sustainability impact assessment
- Sustainability appraisal
- Integrated assessment
- Integrated impact assessment
- Integrated sustainability assessment
- Method
- Approach
- Tool
- Process
- Model
- Implementation
- Social, economic, environment
- Integration
- Integrate
- Integrative

Time frame: 2006-

Part 2: Literature search on integrated sustainability indicator systems for urban development scenario assessment in the context of receiving water bodies for urban storm water

Internet search engines/databases used:

- Google Scholar
- Google

University databases used:

- Web of Science
- ProQuest
- ScienceDirect

Keywords/phrases used:

- Sustainability assessment
- Sustainability assessment indicator systems
- Sustainability
- Indicator system
- Indicator
- Integration
- Integrate
- Integrative
- Multi-criteria approaches
- Multi-criteria
- Social-ecological system
- Social-ecological
- Adaptive governance
- Governance
- Resilience
- Consultation
- Pre-requisites
- Transition
- Urban storm water

Time frame: 2000-

Appendix 2: Core generic criteria for sustainability assessments - from Gibson (2006)

Box 1. Core generic criteria for sustainability assessments

Socio-ecological system integrity
the requirement:
Build human–ecological relations to establish and maintain the long-term integrity of socio-biophysical systems and protect the irreplaceable life support functions upon which human and ecological well-being depends.

illustrative implications:

- need to understand better the complex systemic implications of our own activities;
- need to reduce indirect and overall as well as direct and specific human threats to system integrity and life support viability.

Livelihood sufficiency and opportunity
the requirement:
Ensure that everyone and every community has enough for a decent life and that everyone has opportunities to seek improvements in ways that do not compromise future generations' possibilities for sufficiency and opportunity.

illustrative implications:

- need to ensure provision of key prerequisites for a decent life (which, typically, are not now enjoyed by those who have little or no access to basic resources and essential services, who have few if any satisfactory employment opportunities, who are especially vulnerable to disease, or who face physical or economic insecurity);
- need to appreciate the diversity, and ensure the involvement, of those whose needs are being addressed.

Intragenerational equity
the requirement:
Ensure that sufficiency and effective choices for all are pursued in ways that reduce dangerous gaps in sufficiency and opportunity (and health, security, social recognition, political influence, and so on) between the rich and the poor.

illustrative implications:

- need to build sustainable livelihoods for all, including practically available livelihood choices and the power to choose;
- need to emphasize less materially- and energy-intensive approaches to personal satisfactions among the advantaged, to permit material and energy sufficiency for all.

Intergenerational equity
the requirement:
Favour present options and actions that are most likely to preserve or enhance the opportunities and capabilities of future generations to live sustainably.

illustrative implications:

- need to return current resource exploitation and other pressures on ecological systems and their functions to levels that are safely within the perpetual capacity of those systems to provide resources and services likely to be needed by future generations;
- need to build the integrity of socio-ecological systems, maintaining the diversity, accountability, broad engagement and other qualities required for long-term adaptive adjustment.

Resource maintenance and efficiency
the requirement:
Provide a larger base for ensuring sustainable livelihoods for all, while reducing threats to the long-term integrity of socio-ecological systems by reducing extractive damage, avoiding waste and cutting overall material and energy use per unit of benefit.

illustrative implications:

- need to do more with less (optimize production through decreasing material and energy inputs and cutting waste outputs through product and process redesign throughout product lifecycles), to permit continued economic expansion where it is needed, with associated employment and wealth generation, while reducing demands on resource stocks and pressures on ecosystems;
- need to consider purposes and end uses, recognizing that efficiency gains are of no great value if the savings go to more advantages and more consumption by the already affluent.

Socio-ecological civility and democratic governance
the requirement:
Build the capacity, motivation and habitual inclination of individuals, communities and other collective decision-making bodies to apply sustainability requirements through more open and better informed deliberations, greater attention to fostering reciprocal awareness and collective responsibility, and more integrated use of administrative, market, customary and personal decision-making practices.

illustrative implications:

- need governance structures capable of integrated responses to complex, intertwined and dynamic conditions;
- need to mobilize more participants, mechanisms and motivations, including producers, consumers, investors, lenders, insurers, employees, auditors, reporters
- need to strengthen individual and collective understanding of ecology and community, foster customary civility and ecological responsibility, and build civil capacity for effective involvement in collective decision-making.

Precaution and adaptation
the requirement:
Respect uncertainty, avoid even poorly understood risks of serious or irreversible damage to the foundations for sustainability, plan to learn, design for surprise, and manage for adaptation.

illustrative implications:

- need to act on incomplete but suggestive information where social and ecological systems that are crucial for sustainability are at risk;
- need to design for surprise and adaptation, favouring diversity, flexibility and reversibility;
- need to prefer safe fail over fail-safe technologies;
- need to seek broadly comprehensible options rather than those that are dependent on specialized expertise;
- need to ensure the availability and practicality of back-up alternatives;
- need to establish mechanisms for effective monitoring and response.

Immediate and long term integration
the requirement:
Apply all principles of sustainability at once, seeking mutually supportive benefits and multiple gains.

considerations:

- integration is not the same as balancing;
- because greater efficiency, equity, ecological integrity and civility are all necessary for sustainability, then positive gains in all areas must be achieved;
- what happens in any one area affects what happens in all of the others;
- it is reasonable to expect, but not safe to assume, that positive steps in different areas will be mutually reinforcing.

illustrative implications:

- need positive steps in all areas, at least in general and at least in the long term;
- need to resist convenient immediate compromises unless they clearly promise an eventual gain.

Source: Gibson *et al* (2005)

Appendix 3: Guidance on trade-off rules - from Sadler (2008)

Definitions of sustainability objective:

Following World Bank delimitations of the substitutability of natural capital, three levels of sustainability can be identified that offer a choice of frameworks for evaluating development trends or actions:

43. ***Weak sustainability*** involves maintaining total capital without regard to its composition and allows natural capital to be freely converted into economic capital and output (governed only by existing environmental policies, regulations and guidelines)

44. ***Moderate sustainability*** requires that attention is also given to the mix of capital stocks with natural capital considered substitutable only up to certain critical limits or thresholds (which are not yet known but can be formulated using the precautionary principle)

45. ***Strong sustainability*** means maintaining natural capital more or less at current levels (no net loss) so that losses and damages from development must be replaced or offset in kind (which represent a stringent interpretation of the precautionary and polluter-pays principles)

Guidance for relationship between 'Net effect (gain/loss)' and 'Decision criterion':

Net effect (gain/loss)	Decision criterion
Quadruple win (+ + + +)	Acceptable – gains in all four components
Very positive (+ +)	Acceptable – gains significantly outweigh losses with only non-major impacts in all four components
Positive (+)	Acceptable – gains outweigh losses with no potentially significant adverse impact(s) on any component
Neutral	Not acceptable as is – gains must be enhanced and/or losses reduced to achieve a positive balance as in the previous category
Negative (!)	Normally unacceptable – losses outweigh gains and/or there are potentially significant adverse impacts on one or more components
Very negative (! !)	Unacceptable – losses significantly outweigh gains or do not meet TBL in one or more components
Quadruple loss (! ! ! !)	Unacceptable – losses in all four components

Set of general and enabling guidelines for sustainability assurance:

(1) At all stages of decision making, *priority should be given to options and actions that do the most good*, then to those that do no harm, and finally to those that accept some adverse effects. ... The order of choice, as shown in Table 4, is first to seek ‘quadruple win’ packages that will have lasting benefit, second to secure options that maximise net gains without any major adverse effects, and third to accept options that have more modest net gains and avoid potentially serious adverse impacts.

(2) In principle, *all other configurations of choice would be unacceptable* within a sustainability framework, strictly defined. ... In reality, the process of determining net gains and undertaking the necessary trade-offs will be much more complicated than implied in the schema. At a minimum, preference should be given to the best practicable option that avoids potentially significant adverse impacts after mitigation measures have been taken into account. In some cases, the choice of a ‘least worst’ option for the avoidance of harm may be necessary ...

(3) On some level, *hard choices and trade-offs are an inevitable part of decision making*. This task must be confronted rather than circumvented in conventional practice or assumed away in idealised or normative prescriptions (Toman 1998). It should be undertaken in a transparent and accountable manner. A key means of doing so is to place the onus of proof on the proponent for all trade-offs that assume potentially major or significant adverse effects can be avoided or mitigated. This burden presumes that such effects are unacceptable unless they can be substantiated as reasonably prudent having regard to the circumstances, preferably through an open, rigorous process. Following accepted principles of assessment good practice, the rationale for trade-offs should be subject to explicit justification, e.g. in written reasons for decision.

Appendix 4: Analysis of MCA cases reviewed

Summary of Multi-Criteria Assessment cases reviewed

Author/Case	***A: MCA structural details (1)***				
	Objective	***Alternatives***	***Principles for criteria selection***	***Criteria categories***	***Criteria #***
Ellis et al Daywater (Ch.9) Europe Urban 2008	- to support stakeholder in identifying the most appropriate type of <u>structural</u> BMP for a particular site	- 15 structural BMPs	- should incorporate consideration of environmental, technological, economic and social values - nested hierarchy of criteria and associated KPIs	- selected after DayWater end-user consultation - technical (3 KPIs) - environmental (3 KPIs) - Op & Maintenance (2 KPIs) - Social & urban community benefits (4 KPIs) - Economic costs (2 KPIs) - Legal & Urban Planning (2)	- 6 criteria categories - 16 KPIs
Proctor&Drechsler Australia Upper catchment 2006	- to inform an optimal long term recreation and tourism strategy for the upper catchment	- developed by jury - 5 in total: - BAU - maximise Ecosystem Serv - maximise social outcomes - maximise econ. outcomes - sustainable mix	- nested hierarchy - lowest level = indicators - exhaustive w.r.to overall objective of decision - mutually exclusive - clearly defined	- selected by the jury in their prep.workshop - Eco.Sys.Serv crit. (9) - Soc/Cult.crit. (4) - Econ.crit. (2)	- optimally 7-12 maximum - started with 15 - eliminated some redundant criteria by negotiation
Hermans et al USA Upper catchment 2007	- to frame multi-stakeholder discussions of river management alternatives ''to help local communities balance the longterm cultural, economic, and environmental health of the watershed through active citizen participation''	- developed by Vermont DEC - 4 in total: - maintain channelisation - active restoration of natural - passive restoration of natural - combination	- to reflect the shared vision and the range of values of the whole group - wanted to reduce # because of time limits and constraints of 'conjoint analysis' of individual preferences	- based on spread of expressed river values - economic (6) - social/cultural (6) - environmental (6)	- initially 18 - reduced to 5 'measurable' criteria - economic (2), social/cultural (1) and environmental (2)

Author/Case	***A: MCA structural details (1)***				
	Objective	***Alternatives***	***Principles for criteria selection***	***Criteria categories***	***Criteria #***
Salgado et al Spain Urban water supply 2009	- ??	- nominated by social actors in survey, refined by researchers - initially 20, finally 8 in total: - heighten the dam - use de-salinated water - re-sue waste water - modernise irrigation system - rationalise g/w use - improve eff urban water use - control urban development - do nothing	- no principles stated	- researchers say that actors (on their own) found this difficult; therefore the research team formulated the criteria based on in-depth interviews. grouped under - economic - environmental - social - institutional	- 11 criteria
Soma Norway Coastal development 2010	- not stated for the case study - in the case study, participants were <u>not</u> told/invited to focus on environmental/sustainability issues (research objectives (1) to trial a new combination of MCE/participatory methods, and (2) to evaluate strengths and weaknesses after the trial	- modified Delphi method - alternatives appear to be expressed as sub-objectives; - 4 in total: - extending building activities - conserving cultural sites - conserving areas for outdoor life - conserving biodiversity - alternatives were expressed as maps indicating possible scenarios of land use under each alternative	- nested hierarchy used - at most, 3 levels of criteria, but scores applied only at the lowest level (levels 2 or 3)	- from in-depth interviews with interest groups - modified Delphi method - expand building activities in the coastal zone - expand outdoor life activities in the coastal zone on the shore - expand boating activities in the coastal zone - expand biodiversity in the coastal zone - expand cultural landscapes in the coastal zone	Level 1 (4) Level 2 (16) Level 3 (25)
Russell & Ward NZ Regional catchments 2010	- to establish sustainability thresholds - to evaluate options for a regional water management strategy - to discuss sub-regional variations in water management options - to achieve ‘moderate sustainability’	- 4 options from prior public consultation phase: - continue to improve current - advance environmental protection before ..... - re-configure consents and infrastructure for .... - advance infrastructure with strong requirements for ...	- consideration of different levels of sustainability to aim for - consideration of 4 well beings/4 pillars - consideration of inter-gen and intra-gen equity	- cultural (3) - economic (4) - environmental (6) - processes (6) - social (7)	Initially 21 Finally 26

Author/Case	***A: MCA structural details (1)***				
	Objective	***Alternatives***	***Principles for criteria selection***	***Criteria categories***	***Criteria #***
Taylor & Fletcher Australia Urban stormwater 2005	- to enable stormwater managers to make decisions about the use of stormwater management measures after careful review of financial, social and ecological considerations - comparative sustainability	- structural stormwater mgt alternatives	- incorporate elements of risk assessment into ipact assessment - consider a projects specific, local objectives (e.g. risk of child drowning) as well as broader SD objectives	- 3 categories: - financial - social - ecological	- no details
Hughey et al RIVAS NZ Wild Rivers 2009	- to determine the significance of a river for a specific value - to inform decision makers of the significance of particular values using a consistent approach	- range of potential future applications, including: - alternative recreational values specified, including - - salmonid angling - whitewater kayaking - swimming -	- consistent - transparent - holistic understanding - representative - quantitative - adaptive - standardised - tiered significance - focussed - iterative - impartial - 4 well beings - also refers to SMARTA criteria for indicator selection	- attributes of a river based on considerations such as - - quality - rarity - diversity - representativeness - substitutability - connectivity - use levels - economic benefits - attributes overing the 4 well beings	- attribute clusters (8) - specific attributes (27) - primary attributes (10) to ensure practicality and implementability
Lai et al Australia Urban 2008	[paper not at this level of detail]	[paper not at this level of detail]	[paper not at this level of detail]	[paper not at this level of detail]	[paper not at this level of detail]

Author/Case	***B: MCA structural details (2)***			
	Weighting criteria	***Assessing options***	***Aggregating the criteria***	***Sensitivity analysis/iteration***
Ellis et al Daywater (Ch.9) Europe Urban 2008	- user defined - sum of all weights = 100% - if a criterion is not important at all, set its weight to 0% - refers to a "swing-weighting" method (ODPM, 2003)	- use of performance scores for each criterion (0-10; broader bands; hi-med-lo; ..) - 3 options are presented: - use of default scores - development of scores by the end user - refinement of preferences by developing scores for a short list of BMP options	- multiplicative aggregation: scores by weights	- where various users/stakeholders cannot agree on appropriate weights, take several sets of weights forward for comparison
Proctor&Drechsler Australia Upper catchment 2006	- individual jurors ranked the criteria (most to least important)	- matrix of indicators developed by expert inputs - some indicators quantitative (physical, $, #) - some indicators qualitative (judgement of hi/low)	- used dedicated software ProDecX - ordinal rankings had to be translated to a cardinal scale - for discussion of aggregating algorithms points to Bana e Costa,1990; Gal et al, 1999; Refsgaard, 2006)	- iterating between software-based results and jury deliberations to explore influence on final rank of changing certain weights
Hermans et al USA ??? 2007	- use of 'conjoint analysis' - not explained	- outranking method used - performance of one alternative relative to others, with 3 possible outcomes: - A outranks B - A is outranked by B - neither of the above - refer to 2 families of outranking MCDA methods - ELECTRE (Roy, 1985) and PROMETHEE (Brans and Mareschal, 2005) - also use of maximise or minimise choice for each decision criterion - indicating trade-off preferences	- within PROMETHEE software - not explict	- repeat of 'conjoint analysis' of individual preferences to capture changes in stakeholder preferences that occurred with learning during deliberation
Salgado et al Spain ??? 2009	- no details provided - all done by the research team using NAIADE software	- no details provided - all done by the research team using NAIADE software	- no details provided - all done by the research team using NAIADE software	- not discussed
Soma Norway Coastal development 2010	- weights applied just at the top level of criteria - weights established by survey responses from 16 citizens - before and after the dialogues	- modified eigenvalue method (see Saaty,1994)	- 3 aggregation techniques used in parallel: - Evamix software - Regime software - weighted summation	- refer to 3 aggregation techniques; all gave consistent rankings of options - report was sent to the interest groups for their comment before finalisation

Author/Case	***B: MCA structural details (2)***			
	Weighting criteria	***Assessing options***	***Aggregating the criteria***	***Sensitivity analysis/iteration***
Russell & Ward NZ Regional catchments 2010	- weighting not used in this process, because no aggregation	- first establish QTL and QBL benchmarks for each criterion - score each option on same scale as the QTL and QBL - create a visual sustainability 'profile' illustrating scores for all criteria in relation to their QTL/QBL benchmarks	- aggregation not required by this process	- process would allow for this, but they ran out of time
Taylor & Fletcher Australia Urban stormwater 2005	- should reflect stakeholder views	- no details	- no details	- no details
Hughey et al RIVAS NZ Wild Rivers 2009	- relative contribution of the attribute to the value - essentially a sensitivity analysis (assume equal weightings to begin with)	- determine indicator thresholds - apply thresholds to determine scores	- simple summation - apply criteria to determine spatial level of significance (national/regional/local)	- iterations, varying weights, to examine influence of different judgements and capture group learning
Lai et al Australia Urban 2008	[paper not at this level of detail]	- refers to distinction between MADM (A=attribute) and MODM (O=objective) methods	- refers to a variety of Decision Support Methods: - elementary - single synthesising criterion - outranking - goal/reference point - refers to various software - PROMETHEE, ELECTRE, NAIADE, REGIME,..	[paper not at this level of detail]

Author/case	***C: Stakeholder/decision maker roles***						
	Stakeholder type	***Stakeholder selection***	***Relationship to ultimate decision makers***	***Facilitator***	***Steps in group process***	***Governance arrangements***	***Other aspects***
Ellis et al Daywater (Ch.9) Europe Urban 2008	- 10 scientific partners (from Sweden to Greece) - 15 Core End Users (from Sweden to Greece) - predominantly regional, municipality organisations involved in water infrastructure	- appears to have been determined by composition of participating organisations - dominated by engineers (note emphasis on structural solutions) and urban hydrologists but also with experts in socio-economics and information technology	- for use of the Multi Criteria Comparator tool (MCC), assumes a single decision maker promoting agreement amongst different stakeholders	- use of the MCC tool is mediated via computer	- identifies a 7-step process in use of the MCC tool with "opportunities for stakeholder involvement in 3 (*) - define objective (*) - site screening - identify/describe criteria/KPIs - develop/quantify benchmark utility scores for each KPI(*) - determine relative importance of KPIs - identify list of preferred BMPs - stakeholder discussion for consensus on preferred BMP	- pilot research and development activity - project coordinator to coordinate organisation of parallel tasks facilitate consensus; - project website, including intranet part and file exchange platform to facilitate detailed interactions; - Advisory Steering Board guiding overall project progress at annual meetings; - smaller working meetings to discuss specific bottlenecks in project development	- DayWater's MCC is an entirely computer-based tool
Proctor&Drechsler Australia Upper catchment 2006	- could be citizen's juries - in this case a jury of nat.res.mgers	- for citizens - random or strat. random - in this case, occupational function	- nat.res.mgers had been involved in developing the original strategy which was one of the options assessed	- jury deliberations facilitated	- choose jury - educate jury - jury choose objectives and alternatives to asses - jury select criteria - jury determine weighting of criteria - can use individual weights or seek consensus on weights - jury assess options (no details) - aggregate criteria - computer-assisted - sensitivity analysis, deliberating, iterating	- research case	- jury preparation was important - prior workshop to understand the process and select criteria and options - jury allowed to invite experts

Author/case	***C: Stakeholder/decision maker roles***						
	Stakeholder type	***Stakeholder selection***	***Relationship to ultimate decision makers***	***Facilitator***	***Steps in group process***	***Governance arrangements***	***Other aspects***
Hermans et al USA Upper catchment 2007	- White River Partnership (WRP)- ''a locally led and community driven collaborative between communities, citizens, conservation groups, and federal and state agencies''	- Upper River Stream Team (URST) >> ~20 core members representing local gov't, pvt landowners, farmers, US Forest S; - formed on basis of past interest and active involvement in restoration activities; - direct invitation of WRP staff	- URST acted as both the decision-making body for its area and as a special interest group advocating for external support	- professional facilitator, experienced in environmental facilitation was used in each meeting	- group was educated by scientists about river dynamics, ecology, mgt alternatives; - group collaborated to create future vision and brainstormed important values; - vision and values translated into achievable objectives and 18 criteria; - expert sub-group reduced 18 to 5 measurable criteria later agreed by the whole group; - 'conjoint analysis' to elicit/quantify preferences;	- research on a trial process - WRP formed 6 stream teams around the 6 main sub-watersheds, working within a River Management program proposed by Vermont DoEnvtCons - WRP staff invited URST participants - expert sub-group for consensus on 5 criteria - agreed by whole group - various inputs of experts	- fair and open facilitation important - monthly meetings over 10 months - URST designed and facilitated a public opinion survey to compare their knowledge and values with those of the general public
Salgado et al Spain Urban water supply 2009	- public admin. (5) - businesses (5) - NGOs (4) - experts (2)	- appears to have been by researcher selection, at least initially; followed recent Agenda 21 and Malaga Forum consultation processes - for guidance; - described as 'social actors directly involved in water mgt'	- included those with statutory responsibility for decision making and other not traditionally involved in decision making	- only one focus group; no mention of facilitation.	- diagnose and define the problem; identify other actors - 16 in-depth intervws - questionnaire to actors to identify alternatives and criteria -	- research case	- the Agenda 21 and Malaga Forum activities provided some introductions for the research team to the social actors - Malaga Forum was specifically water-related - seems a research team-centred approach

Author/case	***C: Stakeholder/decision maker roles***						
	Stakeholder type	***Stakeholder selection***	***Relationship to ultimate decision makers***	***Facilitator***	***Steps in group process***	***Governance arrangements***	***Other aspects***
Soma Norway Coastal development 2010	- interest groups - citizens groups (26selected/16 attended) - experts - facilitator	- facilitatory used snowballing technique to identify 'relevant' interest groups; - citizens selected by facilitator (gender, age, location, background) from respondents to letter of invitation sent to randomly recipients; - experts identified by facilitator or called as witnesses by citizens during deliberations	- although not a live decision-making situation, the outputs were used as background information for a new municipality plan, which gave added meaning to the experiment	- in this approach, the 'facilitator' plays a central role with numerous responsibilities covering - - participant selection, - interviewing interest groups, - collating info on alternatives, criteria, scores, ... - summarising and reporting findings - it appears that a separate 'moderator' facilitated the citizen deliberations	- 4 main steps (1) selecting participants (2) Interest groups and experts identifying criteria, alternatives and impacts (3) Citizens judging what is more or less important and in applying scores (4) Summarising and reporting the main findings	- research case	- interest groups involved via in-depth interviews and Delphi surveys - citizens groups involved in 2 face-to-face deliberative sessions + before and after surveys
Russell & Ward NZ Regional catchments 2010	- CWMS Steering Group - CWMS Officials Group	- within the CWMS process	- within the CWMS, these were the decision makers	- workshop processes were facilitated - required prior customisation of information sets - inputs from 'selected specialist experts in social, economic and environmental matters'	- selecting levels of sustainability to reference for trade-off decisions - identifying/prioritising capital assets involved - prepare space-time analyses to record short- and long-term impacts - review/revise evaluation criteria and impact scales - identify QTLs and QBLs - collective scoring - discussion of sub-regional options	- CWMS governance arrangements: - Canterbury Mayoral Forum -	

Author/case	***C: Stakeholder/decision maker roles***						
	Stakeholder type	***Stakeholder selection***	***Relationship to ultimate decision makers***	***Facilitator***	***Steps in group process***	***Governance arrangements***	***Other aspects***
Taylor & Fletcher Australia Urban stormwater 2005	- compared with European Guidelines, Australian Guidelines encourage greater input from non-technical stakeholders	- no details	- implies having the mandate to make recommendations on a preferred option	- not discussed	- # of steps involving stakeholders depends on prior decision about 'level' of assessment appropriate; - could include the following - identify TBL criteria and indicators - determine relative importance/weightings - develop impact matrix - sensitivity analysis to identify preferred option - recommend preferred option to ultimate decision maker - evaluate project performance and provide feedback for improvements	- method comes from Australian TBL assessment guidelines designed for use by local government or major drainage authorities - 'TBL assessment body' convened by the admin.agency	- one unique aspect of this approach (not elaborated in the text) is the provision for three alternative 'levels of assessment' - high, intermediate, basic - corresponding to differing size, complexity and potential impact, and requiring different levels of effort, resources and stakeholder engagement

Author/case	***C: Stakeholder/decision maker roles***						
	Stakeholder type	***Stakeholder selection***	***Relationship to ultimate decision makers***	***Facilitator***	***Steps in group process***	***Governance arrangements***	***Other aspects***
Hughey et al RIVAS NZ Wild Rivers 2009	- scientists - river value experts - recreational user reps with good local knowledge	- two alternatives expressed: (1) national-level expert panel to derive assessment criteria in Part 1 + regional-level expert panel for local knowledge in Part 2 (2) single panel for Parts 1&2	- when implemented by Councils, the Expert Panels will become the decision makers for making the assessments of river significance, while RMA decision makers will incorporate these assessments into planning or resource consent decisions	- not discussed	Part 1: Assessment criteria - (1) define river value categories/river segments (2) identify attributes corresponding to each value (3) select sub-set of primary attributes (4) identify indicators for each primary attribute Part 2: Determination of Significance - (5) Determine indicator thresholds - hi/med/lo importance (6) Apply indicators/thresholds - (a) populate indicator fields with data, and (b) assign 'threshold score' to indicate relative importance (7) Assign weightings (8) Determine river significance	- pilot research and development activity; - initial guidance from national workshop - regular review and agreement between Expert Panel and Reference Group (sign-off at each stage) -	
Lai et al Australia Urban 2008	[paper not at this level of detail]	[paper not at this level of detail]	[paper not at this level of detail]	[paper not at this level of detail]	[paper not at this level of detail]	[paper not at this level of detail]	

Author/Case	***D: Strengths/weaknesses/issues***			
	Strengths	***Weaknesses***	***Issues***	***Responses to issues***
Ellis et al Daywater (Ch.9) Europe Urban 2008	- multiplicative aggregation seen as more discriminating - (once the MCC tool is familiar to users) ease and speed of use allows a degree of ‘experimenting’ with weights to test concerns and options (but only structural options)	- MCC does not include non-structural as feasible alternatives; - is applicable only to individual “stand alone” BMP devices - difficulties in getting stakeholder agreement on benchmarking values, thresholds and performance data for scores	- MCC comparisons are relative, not absolute - quantitative benchmarks and performance indicators for technical, environmental, economic and O&M maybe easier to identify than for social/cultural and administrative/legal criteria	
Beyeler DayWater (Ch.16) Europe Urban 2008		- balancing complexity/simplicity with cognitive capacity - the gap between scientific knowledge and practical knowledge + the different ways of thinking about a real/abstract situation or problem	- appears from a user perspective that the approach to developing the ADSS has been technocratically driven by ‘experts’ in water mgt and information technology, but the development process has not taken non-expert end users along with it and is therefore largely inaccessible	- more direct involvement of end users in the testing and trialling of the ADSS, but their involvement has to start at the beginning - articulating how end users experience problems/questions and letting the ADSS developers hear and respond to these
Proctor&Drechsler Australia Upper catchment 2006	- allows a complex problem to be broken down into workable units - options, criteria and preferences and making them explicit; - applied in a heuristic (iterative with feedback) way, aids jury learning about complex issues; - citizen juries can be used to aggregate multiple individual preference weights through deliberation to achieve consensus.	- in practice, often does not adequately address the facilitation issue of interaction between analyst and decision makers to elicit revised preferences; - with multiple decision makers, does not normally provide clear guidelines on how to analyse or aggregate multiple weights - tends to use averaged weights which may mask extent of different priorities; - in general citizens juries have not been involved in the task of structuring the decision making	(1) if the number of sub-criteria is uneven (e.g. 9 ecol/4soc/2econ), this can lead to an unintended bias	(2) correct for criteria bias by ensuring that the sum of the weights for each main category is the same - i.e. sum of ecol weights = sum of soc weights = sum of econ weights

Author/Case	***D: Strengths/weaknesses/issues***			
	Strengths	***Weaknesses***	***Issues***	***Responses to issues***
Hermans et al USA Upper catchment 2007	Researchers selected PROMETHEE outranking method because - - theory and methodology easier for stakeholders to understand; -more amenable to stakeholder involvement at every stage - changing criteria weights, testing different preference functions - rankings found to be ‘more stable’ - version II is well suited to large groups with diverse opinions - the structuring of deliberations enabled the group to learn and ‘move forward’ - the structured deliberations appeared to focus more on the preferences/weightings for the 5 criteria than on the final outcome	- too many objectives/criteria can overload individual’s thinking and analysis	(1) ignorance and uncertainty amongst stakeholders (2) data imprecision and incompleteness (3) too many separate criteria	(1) workshops with experts (3) use of expert sub-group to distil 18 criteria down to 5
Salgado et al Spain Urban water supply 2009		- difficulty in finding quantitative and/or accurate information on most of the criteria - survey revealed that the “water problem” was not viewed in the same way by the social actors and the general public		- use of fuzzy variables to address lack of quantitative info

Author/Case	***D: Strengths/weaknesses/issues***			
	Strengths	***Weaknesses***	***Issues***	***Responses to issues***
Soma Norway Coastal development 2010	- the involvement interest groups (clarifying different interests) + citizens groups (comparison of interests) is complementary - the method of interest group involvement focussed on explanation of their perspective rather than lobbying for specific solutions - citizens were involved to make comparative values judgements in a long-term perspective - while framing of citizen deliberations favoured social values, the format did not preclude voicing of individual interests - but did tend to de-legitimise them - deliberative processes gave citizens increased understanding of different points of view - the use of maps/visual illustrations of the alternatives aided transparency - the MCE framework assisted in structuring preparations and processes; also in making information transparent (alternatives, criteria, impact scores, weights) pointed to a clear structure for each participatory process - a good combination of agreed facts and social values	- citizens can become overwhelmed by expert contributions in some situations - distracting them from their long-term focus - having experts/interest group reps select criteria risks missing some criteria considered important by individual citizens; - not clear whether this method has dealt appropriately with trade-offs or compromises involving compensation of one factor loss by another factor gain	- legitimacy and transparency are key factors - dichotomies for participants >> private interests vs public interests; present interests vs future interests - a challenge to experts to stay value-neutral - citizens taking long-term views tend to express different priorities than politicians (often the decision makers) who are more inclined to short-term emphases - time constraints on citizen participation? - the balance of influence between citizens and experts/interest groups; how the separation of roles in the process affects this	- legitimacy is increased if the private/public and present/future nexus are addressed - the eigenvalue method of scoring alternatives worked, but others might be equally workable

Author/Case	***D: Strengths/weaknesses/issues***			
	Strengths	***Weaknesses***	***Issues***	***Responses to issues***
Russell & Ward NZ Regional catchments 2010	- participant co-operation - open deliberations	- difficulties with identifying or agreeing scale descriptors for impacts - gaps in data for some criteria	- workshop of 30 people ran out of time to complete tasks; needed follow-up activities by email to complete MCA aspects - only 10 responded - some groups did not want to score particular criteria for the region as a whole, because of sub-regional differences (appropriate spatial scale?)	
Taylor & Fletcher Australia Urban stormwater 2005	- tool to assist urban stormwater managers to make more systematic, informed, holistic, participatory, transparent, multidisciplinary, defendable, socially acceptable, ecologically sustainable and cost-effective decisions (Taylor, 2005a).	- the resources needed to undertake the assessment process; - the complexity that can be generated if many assessment criteria and/or stakeholders are involved; - no guarantee of a ‘sustainable’ outcome.	- for TBL to be effective, you need: - clear organisational vision and objectives that address each element - thorough scoping of alternatives to encourage innovative options - sufficient resources (time, money, skills, info) - organisational culture prepared to invest in structured, transparent, rigorous assessment involving stakeholders with varying skills and opinions	
Hughey et al RIVAS NZ Wild Rivers 2009		- data limitations, particularly quantitative data availability, can restrict the scope of the method		
Lai et al Australia Urban 2008	- simpler MCDA methods can encourage transparency of logic (valued by stakeholders) and provide a useful structure for communicating decisions	- assumption that preference for criteria is assumed to be independent of each other - double-counting problem when chosen criteria are either redundant or non-exhaustive - complex MCDA methods can be perceived by non-experts as “black box” approaches	- general limitations of ‘measuring sustainability’ due to definitional fuzziness, complexity and value judgements	

Appendix 5: Literature identified through internet and database searching

Part 1: Update of the integrated/sustainability assessment review conducted for the Ecologic Foundation in 2005

Adinyira, E. et al 2007, "A Review of Urban Sustainability Assessment Methodologies", in M. Horner, C. Hardcastle, A. Price, & J. Bebbington (eds), International Conference on Whole Life Urban Sustainability and its Assessment, Glasgow
HAVE FULL TEXT
Sustainability has emerged as a planning concept from its beginnings in economics and ecological thinking and has widely been applied to urban development. Urban sustainability is simply described as a desirable state or set of urban conditions that persists overtime. Just as the task of defining sustainability has progressed in response to early economic thinking, so has the task of its assessment. Many urban sustainability assessment methods can be identified from literature. However an examination of these methods reveals largely three methodological foundations. Focusing on the context of urban development, this paper presents an appraisal of the relative potentials and limitations of methods developed around the three identified methodological foundations. The paper agrees with the much held view that, most currently available urban sustainability assessment methods fail to demonstrate sufficient understanding of the interrelations and interdependencies of social, economic and environmental considerations. It further points to a wide gap between assessment theories and practices. To help narrow this rather wide gap, the paper recommends a pragmatic shift in focus, from theory development to application and auditing. A suggestion is made for the application of key assessment methods in a given urban area and across various issues, spatial and time scales so as to allow for method comparison. It is hoped that the parallel application of existing methods will greatly accelerate the urban sustainability assessment learning process and will help in the improvement of both theory and practice

Arbter, K. 2007, "Chapter 5: Strategic Environmental Assessment and Sustainability Impact Assessment: two participatory assessment tools for sustainability" in Evaluating Sustainable Development. Sustainable Development in Europe. Concepts, Evaluation and Applications, Hardi, P. & Martinuzzi, A. (eds), Edward Elgar Publishing Limited, United Kingdom.
FULL CHAPTER APPEARS TO BE AVAILBLE ONLINE AT THE FOLLOWING WEB ADDRESS
http://books.google.co.nz/books?hl=en&lr=&id=gFqS3II1FkYC&oi=fnd&pg=PA83&dq=tool+integrate+OR+integration+OR+integrative+%22integrated+assessment%22+-climate+-medical&ots=EVCUGcxU1T&sig=8J4YWhJUAglVbk37VCtAvjtOrCY#v=onepage&q&f=false

Bond, A. & Morrison-Saunders, A. 2009, "Sustainability appraisal: jack of all trades, master of none?" in Impact Assessment and Project Appraisal, vol. 27, no. 4, pp. 321-329
HAVE FULL TEXT
Sustainable development is a commonly quoted goal for decision making and supports a large number of other discourses. Sustainability appraisal has a stated goal of supporting decision making for sustainable development. We suggest that the inherent flexibility of sustainability appraisal facilitates outcomes that often do not adhere to the three goals enshrined in most definitions of sustainable development: economic growth, environmental protection and enhancement, and the wellbeing of the human population. Current practice is for sustainable development to be disenfranchised through the interpretation of sustainability, whereby the best alternative is good enough even when unsustainable. Practitioners must carefully and transparently review the frameworks applied during sustainability appraisal to ensure that outcomes will meet the three goals, rather than focusing on a discourse that emphasises one or more goals at the expense of the other(s)

Bond, A. & Morrison-Saunders, A. 2010, "Re-evaluating Sustainability Assessment: Aligning the vision and the practice" in Environmental Impact Assessment Review, In Press
HAVE FULL TEXT
Sustainable Development is the core goal of the expanding field of Sustainability Assessment (SA). However, we find that three key areas of debate in relation to SA practice in England and Western Australia can be classified as policy controversies. Through literature review and analysis of documentary evidence we consider the problem of reductionism (breaking down complex processes to simple terms or component parts) as opposed to holism (considering systems as wholes); the issue of contested understandings of the meaning of sustainability (and of the purpose of SA); and the

definition of 'inter-generational' in the context of sustainable development and how this is reflected in the timescales considered in SA. We argue that SA practice is based on particular framings of the policy controversies and that the critical role of SA in facilitating deliberation over these controversies needs to be recognised if there is to be a move towards a new deliberative sustainability discourse which can accommodate these different framings

Boulanger, P. Integration in Sustainability Assessment: Meanings, Patterns and Tools, Draft, http://users.skynet.be/idd/documents/EIDDD/WP02.pdf
HAVE FULL TEXT

Cardwell, H. et al.2006, "Integrated Water Resources Management: Definitions and Conceptual Musings", in Journal of Contemporary Water Research & Education, vol. 135, issue 1, pp. 8-18
HAVE FULL TEXT
Integrated Water Resources Management (IWRM) is widely endorsed over the alternative, yet a deeper look at what the term means and what it implies has merits This paper derives a working definition that emphasizes a unified process directed toward achievement of a common goal. An analysis of U.S. Federal policy, specifically that of the Army Corps of Engineers' Civil Works program, suggests that the common goal for Federal water resources management is sustainable development. We present a framework, using the axes of time, space, institutions, and objectives, for examining the nature and degree of management integration. Finally, we compare this simple derivation to definitions of IWRM from various institutions, and touch upon the relationship of the term "watershed approach" to IWRM as used in U. S. Federal agencies. Increasing demands on water resources from growing populations, increased concern for environmental quality, and greater recognition of the interrelationships of competing water resource demands prompted the call for a more comprehensive, coordinated, unified and integrated approach to water resource problem-solving. This approach requires consideration of the interactions among different natural resource elements, such as ecology, hydrology, and geomorphology. among different disciplines, such as economics, engineering, and biology, and among different institutions, including federal, state, and nongovernmental. In the most recent U.S. Army Corps of Engineers Civil Works Strategic Plan, this approach is called Integrated Water Resources Management (IWRM). It draws on a long history of conceptual development. Water resources are often public resources in the U.S. where government agencies have long dominated their management. Laws authorizing and regulating water use and management have proliferated due to growing demands on existing supplies. Agency compliance inevitably forces more interaction with other agencies and nongovernmental organizations (NGOs), if not more integration of action. Many obstacles impede more effective interaction. Increasing appreciation for the complexity of water resource management systems has been accompanied by increased recognition of inter-organizational collaboration, beyond the minimum required by law, as the key to more effective management, which is increasingly defined by a sustainable development goal. With a broadly acceptable goal held in common, this thesis argues that more integrated, collaborative approaches to water resources management will result in more sustainable water resources development because they more completely reflect societal values and scientific knowledge, and focus them on solving complex management problems in a more comprehensively satisfying way. Because future progress depends in part on common understanding of the concept, we revisit different definitions of IWRM and propose a simple conceptual framework for consideration.

Cavanagh, A. et al. 2006, "The Sustainability Assessment Model (SAM): Measuring Sustainable Development Performance", in Australasian Journal of Environmental Management, vol. 13, pp. 142-145
HAVE FULL TEXT
Increasingly, legislative requirements are being placed on councils and businesses to incorporate sustainability into their activities. However, it is inherently difficult to assess the sustainability of activities and there are few tools to do this. It is therefore difficult to understand if an individual project contributes to broad sustainability objectives or if it incorporates understanding of sustainability into organisations' activities. The Sustainability Assessment Model (SAM) is a tool for engaging people within organisations in sustainable development thinking and to evaluate the sustainability of projects (Baxter et al. 2002; Bebbington and Frame 2003). This article provides an overview of the SAM and presents a preliminary assessment of how it was used to assess organic waste processing options for a local council in New Zealand.

De Ridder, W. et al. 2007, 'A framework for tool selection and use in integrated assessment for sustainable assessment' in Journal of Environmental Assessment Policy & Management, vol. 9, issue 4, pp. 423—441
[In Tools, Techniques & Approaches for Sustainability. Collected Writings in Environmental Assessment Policy and Management, Sheate, W. (ed), World Scientific Publishing Co, Singapore.]
BOOK – SOME TEXT ONLINE
http://books.google.co.nz/books?id=m3nCSDxWXUUC&pg=PA26&lpg=PA26&dq=Gibson,+R.+2006.+%E2%80%9CBeyond+the+Pillars:+Sustainability+Assessment+as+a+Framework+for+Effective+Integration+of&source=bl&ots=TZOk7DTD5h&sig=ttEN82YcjfbqdW91qAEyEGh1y2Y&hl=en&ei=b8kaTNLcIZGENr74-dwF&sa=X&oi=book_result&ct=result&resnum=8&ved=0CC4Q6AEwBw#v=onepage&q=Gibson%2C%20R.%202006.%20%E2%80%9CBeyond%20the%20Pillars%3A%20Sustainability%20Assessment%20as%20a%20Framework%20for%20Effective%20Integration%20of&f=false

Dovers, S. 2005, "Clarifying the Imperative of Integration Research for Sustainable Environmental Management" in Journal of Research Practice, vol. 1, issue 2
HAVE FULL TEXT
This paper discusses why integration is important in doing research for developing policy and practice of sustainable environmental management. The imperative of integration includes environmental, social, economic, and other disciplinary considerations, as well as stakeholder interests. However, what is meant by integration is not always clear. While the imperative is being increasingly enunciated, the challenges it presents are difficult and indicate a long term pursuit. This paper clarifies the different dimensions of integration, as an important preliminary step toward advancing mutual understanding and the development of approaches. The paper identifies the driving forces for integration, discusses when integration is required, categorises forms of integration, and proposes principles to inform research programs and projects

Eales, R. et al, 2005, "Integrated appraisal" in Impact Assessment and Project Appraisal, vol. 23, no. 2, pp. 113-123
HAVE FULL TEXT
This paper reviews trends in integrated appraisal with specific emphasis on work within UK Government. It shows how these newer approaches link with current appraisal methods. A two-stage approach is increasingly advocated: first, the potential impacts of the proposal under consideration are 'screened' against a range of economic, social and environmental criteria; second, there is more detailed appraisal where this is considered necessary using appropriate appraisal tools. A brief review is given of existing appraisal tools and their relationships to one another: appraisal should facilitate the use of different tools for different purposes. Finally, we consider the challenges and opportunities of integrated appraisal

El-Haram M. et al. 2007, "Development of an Integrated Sustainability Assessment", Paper submitted to International Conference on Whole Life Urban Sustainability and its Assessment, Glasgow
HAVE FULL TEXT
Integration of environmental, social and economic issues is still a key challenge for the delivery of sustainable development and accordingly sustainability assessment. In attempting to address this challenge in relation to the built environment, SUE-MoT (aconsortium of academic, public and private partners) has been formed with a particular focus on developing holistic metrics, methods and tools in a way that reflects stakeholder values. Currently SUE-MoT is working to develop an Integrated Sustainability Assessment Toolkit (ISAT) that brings together many approaches, allowing key decision-makers to identify the most appropriate for their project and to combine the results based on their values. The objective of this paper is to report on the principles of developing the ISAT so that it will allow a range of key decision makers to systematically and transparently make holistic sustainability assessment decisions for a variety of projects. This paper begins by defining the principle of sustainability assessment and describing the ISAT framework process, highlighting the major stages and the general tasks involved at each stage. The ISAT's major structures including the spatial scale of assessment, the life cycle, sustainability issues and their impact, the sustainability assessment tools database, stakeholder values and finally the mechanisms for stakeholder engagement throughout the process of assessment are also described. The sustainability impact assessment methodologies (based on principles of environmental impact assessment, social impact assessment and cost and benefits analysis) are briefly discussed

Federal Office for Spatial Development ARE. 2004, Sustainability assessment: Conceptual framework and basic methodology, www.are.ch
HAVE FULL TEXT

Garmendia, E. & Stagl, S. 2010, "Public participation for sustainability and social learning: Concepts and lessons from three case studies in Europe" in Ecological Economics, vol. 69, issue 8, pp. 1712-1722
HAVE FULL TEXT
Shaping change such that it avoids losing potentially useful options for future development is a challenging task in the face of complex, coevolving socio-ecological systems. Sustainability appraisal methods, which open up dialogue and options before closing down and making suggestions, pay attention to the inclusion of various and conflicting points of view and address uncertainty, are increasingly used in the science, environment and energy policy domains. The quality of the process is seen as key to high quality appraisal outcomes. Dimensions of quality include learning opportunities which are seen as ways for addressing complexity and uncertainty. Participatory sustainability appraisal methods intend to support social learning among participants. Despite high expectations, social learning processes in sustainability appraisals are poorly conceptualized and empirically understudied. This paper (1) briefly reviews theories of social learning; (2) develops a conceptual framework for the analysis; and (3) presents an empirical application of the framework by use of data obtained from three energy and natural resource management case studies around Europe

Gasparatos, A. et al 2009, "The argument against a reductionist approach for measuring sustainable development performance and the need for methodological pluralism" in Accounting Forum, vol. 33, pp. 245-256
HAVE FULL TEXT
Both sustainability and sustainable development continue to remain elusive concepts even now, 20 years after the Brundtland Commission report that brought them into prominence. This situation most likely stems from the fact that sustainability science encompasses the need to address a wide set of issues over different time and spatial scales and thus inevitably accommodates opinions from diverse branches of knowledge and expertise. However, despite this multitude of perspectives, progress towards sustainability is usually assessed through the development and utilisation of single sustainability metrics such as monetary tools, composite sustainability indices and biophysical metrics including emergy, exergy and the ecological footprint. But is it really justifiable to assess the progress towards sustainability by using single metrics? This paper argues that such a choice seems increasingly unjustifiable not least due to these metrics' methodological imperfections and limits. Additionally, our recent awareness of economies, societies and ecosystems as complex adaptive systems that cannot be fully captured through a single perspective further adds to the argument. Failure to describe these systems in a holistic manner through the synthesis of their different non-reducible and perfectly legitimate perspectives amounts to reductionism. An implication of the above is the fact that not a single sustainability metric at the moment can claim to comprehensively assess sustainability. In the light of these findings this paper proposes that the further elaboration and refinement of current metrics is unlikely to produce a framework for assessing the progress towards sustainability with a single metric. Adoption of a diverse set of metrics seems more likely to be the key for more robust sustainability assessments. This methodological pluralism coupled with stakeholder involvement seems to offer a better chance of improving the outcome of the decision making process.

Gasparatos, A. 2010, "Embedded value systems in sustainability assessment tools and their implications" in Journal of Environmental Management, vol. 91, pp. 1613-1622
HAVE FULL TEXT

Gibson, R. 2006, "Sustainability assessment: basic components of a practical approach" in Impact Assessment and Project Appraisal, vol. 24, no. 3, pp. 170-183
HAVE FULL TEXT
The last few years have brought many experiments with forms of sustainability assessment, applied at the strategic and project levels by governments, private-sector firms, civil society organizations and various combinations. The attractiveness of the work so far suggests that it is now time to prepare for comprehensive adoption and more consistent application of the requirements and processes. The key first steps in sustainability assessment regime design are addressed in this paper. They centre on the

basic sustainability requirements that should inform a transition to sustainability assessment; the main implications of these requirements for sustainability assessment decision criteria and trade-off rules; how to incorporate proper attention to the specific circumstances of applications into particular cases and contexts; and, more generally, how to design practical sustainability assessment regimes.

Gibson, R. 2006. “Beyond the Pillars: Sustainability Assessment as a Framework for Effective Integration of Social, Economic and Ecological Considerations in Significant Decision-Making”, Journal of Environmental Assessment Policy and Management, vol. 8, no. 3, pp. 259-280.
HAVE ABSTRACT James wants; can’t find asked Cawthron
Sustainability is an essentially integrative concept. It seems reasonable, then, to design sustainability assessment as an essentially integrative process and framework for decision-making on undertakings that may have lasting effects. The realm of sustainability has often been depicted as the intersection of social, economic and ecological interests and initiatives. Accordingly, many approaches to sustainability oriented assessments â?” at the project as well as strategic level â?” have begun by addressing the social, economic and ecological considerations separately and have then struggled with how to integrate the separate findings. The problem is exacerbated by the generally separate training of experts in the three fields, the habitual collection of data separately under the three categories and the common division of government mandates into separate social, economic and ecological bodies. The combined effect is not merely an absence of integrative expertise, data and authority but an entrenched tendency to neglect the interdependence of these factors. The three pillars or triple bottom line approach also appears to encourage an emphasis on balancing and making trade-offs, which may often be necessary but which should always be the last resort, not the assumed task, in sustainability assessment. There are, however, important concerns underlying advocacy and application of some three pillar, limited integration approaches. Most significant are well-grounded fears that integrated, sustainability-based assessments may facilitate continued or even renewed neglect of traditionally under valued considerations, especially the protection of ecological systems and functions. This problem needs to be addressed thoughtfully in judgements about how integration is to be done. One possible solution is to take sustainability as an essentially integrative concept and to design sustainability assessment more aggressively as an integrative process. This would entail a package of regime and process design features, centred on ones that.â?¢ build sustainability assessment into a larger overall governance regime that is designed to respect interconnections among issues, objectives, actions and effects, though the full interrelated set of activities from broad agenda setting to results monitoring and response;.â?¢ design assessment processes with an iterative conception-to-resurrection agenda, aiming to maximise multiple, reinforcing net benefits through selection, design and adaptive implementation of the most desirable option for every significant strategic or project level undertaking;.â?¢ redefine the driving objectives and consequent evaluation and decision criteria to avoid the three conventional categories, to ensure attention to usually neglected sustainability requirements and to focus attention on the achievement of multiple, mutually reinforcing gains;.â?¢ establish explicit basic rules that discourage trade-offs to the extent possible while guiding the decision-making on those that are unavoidable;.â?¢ provide means of combining, specifying and complementing these generic criteria and trade-off rules with attention to case- and context-specific concerns, objectives, priorities and possibilities;.â?¢ provide integrative, sustainability-centred guidance, methods and tools to help meet the key practical demands of assessment work, including identifying key cross-cutting issues and linkages among factors, judging the significance of predicted effects, and weighing overall options and implications; and.â?¢ ensure that the decision-making process facilitates public scrutiny and encourages effective public participation.

Graymore, M. et al. 2008, “Regional sustainability: How useful are current tools of sustainability assessment at the regional scale?” in Ecological Economics, vol. 67, issue 3, pp. 362-372
HAVE FULL TEXT
Sustainability assessment methods are primarily aimed at global, national or state scales. However, modelling sustainability at finer spatial scales, such as the region, is essential for understanding and achieving sustainability. Regions are emerging as an essential focus for sustainability researchers, natural resource managers and strategic planners working to develop and implement sustainability goals. This paper evaluates the effectiveness of current sustainability assessment methods - ecological footprint, wellbeing assessment, ecosystem health assessment, quality of life and natural resource availability - at the regional scale. Each of these assessment methods are tested using South East Queensland (SEQ) as a case study. It was selected because of its ecological and demographic diversity, its combination of coastal and land management issues, and its urban metropolitan and rural

farm and non-farm communities. The applicability of each of these methods to regional assessment was examined using an evaluation criteria matrix, which describes the attributes of an effective method and the characteristics that make these methods useful for regional management and building community capacity to progress sustainability. We found that the methods tested failed to effectively measure progress toward sustainability at the regional scale, demonstrating the need for a new method for assessing regional sustainability

Hacking, T. & Guthrie, P. 2008, "A framework for clarifying the meaning of triple-bottom-line: Integrated, and sustainability assessment", in Environmental Assessment Review, vol 28, issue 2-3, pp. 73-89
HAVE FULL TEXT
"Terms such as Integrated Assessment and Sustainability Assessment are used to label 'new' approaches to impact assessment that are designed to direct planning and decision-making towards sustainable development (SD). Established assessment techniques, such as EIA and SEA, are also widely promoted as SD 'tools'. This paper presents the findings of a literature review undertaken to identify the features that are typically promoted for improving the SD-directedness of assessments. A framework is developed which reconciles the broad range of emerging approaches and tackles the inconsistent use of terminology. The framework comprises a three-dimensional space defined by the following axes: the comprehensiveness of the SD coverage; the degree of 'integration' of the techniques and themes; and the extent to which a strategic perspective is adopted. By applying the framework, assessment approaches can be positioned relative to one another, enabling comparison on the basis of substance rather than semantics."

Hayward, L. et al. 2009, "Developing a sustainability assessment framework for planning for sustainability for biofuel production at the policy, programme or project level", Brebbia, C. & Mammoli, A. (eds) Energy and Sustainability II, vol. 121, pp. 355-365…conference?
HAVE ABSTRACT
Biofuel development projects are proceeding at a pace that outstrips normal planning and feasibility evaluation. We present a theoretical framework for planning for sustainability for biofuel production at the policy, plan/programme or project (PPP) level. The objective of this framework is to foster and preserve the social ecological system in which the PPP is to occur so that the system remains dynamic, adaptive, resilient and durable through time. The overall approach to the framework is the understanding of the function and relationships within the social ecological system in which the proposed biofuel PPP is to occur. The system is analysed through stakeholder engagement and expert analysis. Sustainability is assessed through the development of sustainability principles and criteria with sustainability indicators to measure the progression. The framework has developed through a long history of impact assessment and strategic environmental assessment in the environmental sector

Helming, K., et al (eds). 2007, Sustainability Impact Assessment of Land Use Changes, Sringer, Berlin.
BOOK – NOT AVAILABLE AT UNIVERSITY, SOME TEXT ONLINE
Looks interesting – excerpts from introduction - "the need for real integration of economic, environmental and social issues…this book describes results achieved halfway through the SENSOR project…methods for valuing these impacts and integrating them into the wider sustainability context had to be developed. This was achieved through an integration of top-down quantitative modelling and indicator analysis with bottom-up participatory research. The intention of this book is to provide an overview of the various analytical components and their role in the development of sustainability impact assessment tools for multifunctional land use"
http://books.google.co.nz/books?hl=en&lr=&id=5e6NtO5PAMgC&oi=fnd&pg=PA1&dq=approach+integrate+OR+integration+OR+integrative+%22integrated+assessment%22+-climate&ots=IeBh8TqaeE&sig=39uMc3Ww6cDNl0GdDJFgbOBANcA#v=onepage&q=approach%20integrate%20OR%20integration%20OR%20integrative%20%22integrated%20assessment%22%20-climate&f=false

Jacob, K. and Hertin, J. 2007, Evaluating Integrated Impact Assessments – a conceptual framework, EPIGOV Paper No. 7, Ecologic – Institute for International and European Environmental Policy: Berlin.
HAVE FULL TEXT

Impact Assessment (IA) has become a standard procedure of policy making in the European Commission as well as on the national level. It aims to rationalise the decision making process by systematically collecting information about the likely impacts of a planned policy, by evaluating these impacts and thereby providing the basis for choosing the ‘best’ policy option. IA is seen by many as acore instrument in the context of the better regulation agenda which aims to reduce regulatory burden on business. Recently, additional functions have been assigned to this tool, e.g. increasing the transparency of decision making, improving interdepartmental cooperation, and - crucially - addressing crosscutting issues such as environmental policy integration. With these different functions, the institutional design of IA systems has diversified. IA systems can be distinguished regarding a number of features of their institutional framework and practice, for example their timing in decision making process, the degree of transparency and openness for participation, the presence of quality control mechanisms, and the degree to which impacts are quantified. This paper – which draws on the ongoing research in the EVIA project - aims to set out a framework for exploring the contribution of IA to policy integration. This is done by firstly developing a typology of IA systems based on their institutionalisation, orientation and practice and secondly by developing a framework for the analysis of integrated IA, including explicit hypotheses about how different factors are expected to impact on the ability of IA to improve the evidence-base of policy and to promote (environmental) policy integration

Kessler, J. & Abaza, H. 2006, “United Nations Environment Programme’s approach to integrated assessment of trade-related policies: evolution and recent progress” in Impact Assessment and Policy Appraisal, vol. 24, no. 4, pp. 273-283
HAVE FULL TEXT
Over the last decade the Economics and Trade Branch (ETB) of the United Nations Environment Programme (UNEP) has worked to enhance national capacities to undertake integrated assessment of trade-related policies. This article describes the evolution of the approach adopted by UNEP towards achieving greater integration on three key levels: `substantively' by using quantitative and qualitative tools that generate insight into the interrelationships among environmental, social and economic impacts of public policies; `policy-wise' by establishing more direct linkages between the planning and decision-making processes; and `process-wise' by promoting inter-ministerial co-ordination and multi-stakeholder participation and involvement. There is also a shift towards taking a more proactive approach to designing trade policies that will help reduce poverty in an environmentally sound way.

Lee N, 2006, “Bridging the gap between theory and practice in integrated assessment”, Environmental Impact Assessment Review vol. 26, issue 1, pp. 57-78.
HAVE FULL TEXT
There is growing support for the use of integrated assessments (IAs)/sustainability impact assessments (SIAs), at different government levels and geographic scales of policy-making and planning, both nationally and internationally. However, delivering good quality IAs/SIAs, in the near future, could be challenging. This paper mainly focuses upon one area of concern, differences between research and other technical contributions intended to strengthen assessment methodologies and the types of assessment methods considered usable by practitioners. To help in addressing this concern, the development of a common assessment framework is proposed, which is based on a shared, practitioner–researcher–stakeholder understanding of what constitutes a satisfactory integrated/sustainability impact assessment. The paper outlines a possible structure for this framework, which contains three interconnected elements—the planning context in which the assessment is to be carried out; the process by which the assessment is to be undertaken and its findings used; and the methods, technical and consultative, by which impacts are to be assessed. It concludes with suggested ‘next steps’, addressed to researchers, practitioners and other stakeholders, by which the assessment framework might be tested and improved, and its subsequent use supported.

Lotze-Campen, H. 2008, “The role of modelling tools in Integrated Sustainability Assessment (ISA)” in International Journal of Innovation and Sustainable Development, Vol. 3, Nos. 1–2, pp.70–92.
HAVE FULL TEXT
The paper provides a review of existing modelling tools that have been applied to a wide range of sustainability issues and have been used for policy-relevant sustainability assessments. The review covers biophysical models, socio-economic models as well as integrated approaches. It is discussed

at which stages of the Integrated Sustainability Assessment (ISA) cycle models can play an important role. A case study is provided on how three prominent models are to be linked for ISA. The paper closes with a discussion of challenges and limitations in using models in an interdisciplinary setting

Marsden, G. et al. 2010, "Sustainability Assessment: The Definition Deficit" in International Journal of Sustainable Transportation, vol. 4, issue 4, pp. 189-211
HAVE ABSTRACT
Much work has focused on the development of indicator sets to monitor changes in the sustainability of transport. Such indicator sets are however, often quite divorced from those used in decision-making and fail to include clear sustainability goals to work towards. This research describes the development of a sustainability appraisal framework in conjunction with a series of key decision-makers in England. A case study of a real set of strategy options tested in a metropolitan area is outlined and the results used to assess the extent to which current strategy development in the United Kingdom produces the information required to both assess and communicate progress towards sustainability. The results suggest that although sustainability exists as a concept, it is poorly defined. This definition deficit has serious implications for the types of strategies tested. First, information on some aspects of sustainability is not produced and so these aspects are marginalized. Secondly, the lack of policy goals and the dominant welfare economics assessment paradigm allow unsustainable strategies to be justified provided they perform better than an unsustainable 'do-minimum'. The paper concludes with some recommendations for the policy and research communities to bridge the current gap in thinking

Mathur, V. 2007, "Defining, identifying and mapping stakeholders in the assessment of urban sustainability", M. Horner, C. Hardcastle, A. Price, & J. Bebbington (eds), International Conference on Whole Life Urban Sustainability and its Assessment, Glasgow
HAVE FULL TEXT
The assessment of sustainability requires that the diverse values of the stakeholders are represented in the context-specific interpretation of sustainability and in the choice of a desired course of action. Sustainability is a broad concept, and the stakeholders in sustainability are many. In order to have effective stakeholder engagement, it is crucial that all the relevant stakeholders are identified early in the process. In urban development projects, some stakeholders may be obvious, but there might be others who are excluded from the usual decision-making processes and these may even bear disproportionate environmental, social or economic costs leading to inequitable outcomes. This has created the need for a systematic approach to defining and identifying stakeholders for different contexts. This paper will evaluate existing approaches for defining and identifying stakeholders in development projects and the requirements of a sustainability assessment process. Through this analysis, it will identify/develop an approach for defining and identifying stakeholders that is most appropriate for sustainability assessment. The paper also argues that it is important to map out the levels of interest of different stakeholders in relation to the power that they hold. This is useful in determining the appropriate engagement techniques at the different stages of a project and also in understanding any potential conflicts. It is thus important to understand the relationships between the different stakeholders because this can affect the success of the engagement process. Such a mapping of stakeholders can also be useful in anticipating the expectations of the different stakeholders from the project

Morrison-Saunders, A. & Therival, R. 2006, "Sustainability Integration and Assessment", in Journal of Environmental Assessment Policy and Management, vol. 8, no. 3, pp. 281-298
HAVE FULL TEXT
The broadening out of environmental assessment to also consider social and economic dimensions poses some unique challenges, not the least of which is understanding exactly what such a process might entail. This paper outlines the spectrum of possibilities and explores the issue of when and how environmental, social and economic considerations can be integrated in sustainability assessment. The integration issue is also relevant to the practice of strategic environmental assessments (SEA).A new way of conceptualising these types of assessment is put forward based on: (i) what is being assessed — the "question" that is being asked; and (ii) what approach is being used—the type of assessment selected from the spectrum of possibilities. The latter ranges from impact minimisation for each of the three sustainability pillars through to sustainability considered as an integrated concept. The combination of the question and assessment approach determines the level, extent and timing of integration of environmental, social and economic considerations that can be achieved. Additional thought needs to be given to who is performing the integration role as well as the nature of a particular

proposal or its setting. This approach to thinking about SEA and sustainability assessment is illustrated with examples from Australia and the United Kingdom.

Ness, B. et al 2007, “Categorising tools for sustainability assessment” in Ecological Economics, vol. 60, pp. 498-508
HAVE FULL TEXT
The aim of this paper is to provide a categorisation of sustainability assessment tools within the broader objective of lifting the understanding of sustainability assessment from the environmental-focused realm to a wider interpretation of sustainability. The suggested framework is based on three main categories: indicators/indices, product-related assessment, and integrated assessment tools. There is furthermore the overarching category of monetary valuation tools that can be used as a part of many of the tools listed in the three categories. The tools are also divided by their spatial focus and the level of nature–society system integration. Discussion focuses on if and how the tools fulfil the objectives from the more current understanding of sustainability assessment

Nilsson, M. 2006, Sustainability Appraisal Tools, Online. URL: http://ivm5.ivm.vu.nl/sat/chapdb.php?id=106.
CAN’T FIND

Office of the Deputy Prime Minister. 2005, Sustainability Appraisal of Regional Spatial Strategies and Local Development Documents. Guidance for Regional Planning Bodies and Local Planning Authorities, London, UK, Office of the Deputy Prime Minister Publishing.
HAVE FULL TEXT

Pope, J., Morrison-Saunders, A., & Annandale, D. 2005, “Applying sustainability assessment models” in Impact Assessment and Project Appraisal, vol. 23, no. 4, pp. 293-302.
HAVE FULL TEXT
Three sustainability assessment models are applied to a major Australian resource proposal — the Gorgon gas development. `Environmental impact assessment (EIA)-driven integrated assessment' resulted in significant environmental resources being `traded' for socio-economic benefits. `Objectives-led integrated assessment' seeks to maximise social, economic and environmental objectives set by decision-makers. The Gorgon assessment focused on meeting the proponent's strategic objectives, thus missing an opportunity to maximise benefits for the wider community. `Assessment for sustainability' uses sustainability criteria determined by society. The Western Australian Government has recently begun to derive such criteria and the Gorgon proposal would have failed to meet some of them. The actual Gorgon assessment was conducted within a sustainability framework, but the EIA-based approach used did not result in sustainable outcomes. An `assessment for sustainability' approach offers the most promising avenue for future applications.

Pope, J. 2006b, “Editorial: What’s so special about sustainability assessment?” in Journal of Environmental Assessment, Policy and Management, 8(3), v-ix
CAN’T FIND

Pope, J. 2007, Facing the Gorgon: Sustainability assessment and policy learning in Western Australia, Unpublished PhD thesis. Murdoch University
HAVE FULL TEXT
Sustainability assessment is emerging as a form of impact assessment with the concept of sustainability at its heart. This thesis contributes to the process of theory-building for sustainability assessment through an exploration of the development of this policy tool within the Western Australian context. Through an analysis of the sustainability assessments of the Gorgon gas development on Barrow Island and the South West Yarragadee water supply development, and a process of personal reflection, I explore the potential of sustainability assessment to contribute to a more sustainable society by facilitating learning. While the focus of traditional forms of impact assessment has typically been ‘exterior’ forms of knowledge and learning relating to the potential impacts of a proposal, or to process methodologies and governance arrangements, in this thesis I argue that sustainability assessment processes should also facilitate ‘interior’ forms of learning that excavate and challenge underpinning assumptions about the organisation of society, including shared discourses and storylines, as well as personal views and beliefs. To achieve this aim, I maintain that sustainability

assessment should be a proactive process that is integrated with the proposal development, framed by an open question and guided by a ‘sustainability decision-making protocol’ that operationalises sustainability for the decision at hand. It should be guided by a structured process framework that assures attention is given to issues that might otherwise be neglected. Each step of the process framework should represent a space for inclusive deliberation, with the concept of sustainability itself acting as a catalyst for learning and reflexivity.
Located within the institutions of modern industrial society, deliberative sustainability assessment processes can contribute to the emergence of an 'integral sustainability' that embraces and reconnects the interior and exterior, collective and individual dimensions of policy-making and of society in general. The influence of sustainability assessment can thus extend beyond the immediate decision at hand to contribute to a momentum for societal change towards a more sustainable future

Pope, J. 2007, Sustainability assessment of strategic projects: Reflections from Western Australia, Paper presented at Simposio de Desarrollo Sustentable del Sector Eléctrico (Symposium on sustainable development in the electrical sector), Mexico City October 2007
CAN’T FIND

Pope, J. 2007, Sustainability assessment as a deliberative learning process Paper presented at the 3rd International Conference on Environmental, Cultural, Economic and Social Sustainability. Chennai, India, January 2007.
HAVE ABSTRACT James wants; can’t find; has emailed Jenny directly
Sustainability assessment is emerging internationally as the 'third generation' of impact assessment processes, following project environmental impact assessment (EIA) and the strategic environmental assessment (SEA) of policies, plans and projects. However, there is little consensus as to how sustainability assessment should be conducted or what role it might play within a governance structure for sustainability. While sustainability assessment is often conceptualised as a process whose aim is to predict and enhance the environmental, social and economic impacts of a proposal, this paper argues that its transformative potential extends far beyond this. Drawing upon two recent project case studies from Western Australia, it discusses how sustainability assessment can provide a forum for reflexive deliberation and be a powerful vehicle for social learning and change.
Pope’s email: http://integral-sustainability.net/contact-us/

Rotmans, J., Jager, J. & Weaver, P. 2008, “Editorial” in International Journal of Innovation and Sustainable Development, vol. 3, no. 1/2
HAVE FULL TEXT, EXCERPT:
The essential design requirements for ISA stem directly from its intended role as a process for exploring and supporting issue reframing, policy reorientation (regime change) and approaches to transition. ISA represents a new mode of knowledge production that responds to the governance and management challenges of sustainable development. It offers a forum for defining ‘socially- and ecologically-robust’ targets and thresholds concerning conditions to attain or to avoid, integrating these as elements of operational, context-specific sustainability interpretations and exploring alternative pathways of transition. Methodologically, ISA combines three elements: an integrated systems analysis, which seeks to secure broad scope for the assessment; a multi-level and agent-based analytical approach, which seeks to understand multi-level processes that could lead to structural change; and a cyclical, participatory process-architecture, which seeks to promote social learning among stakeholders through dialogue, experimentation, and capacity-building. To handle the complexity and concerns of sustainable development, ISA employs scale- and domain-exceeding concepts, such as stocks, flows and agents, and uses multiple time horizons that may extend over generations.

Sadler, B., Ward, M. & Frame, B. 2008, A Framework for Sustainability Appraisal in New Zealand, Landcare Research Contract Report: LC0708/090
HAVE FULL TEXT

Sadler, B. 2010, “Toward Sustainability Appraisal: Frameworks, Approaches, and Tools” CAN’T SEEM TO DOWNLOAD, CAN VIEW AT THIS ADDRESS
http://webcache.googleusercontent.com/search?q=cache:Sx_XGjhKVoJ:www.csinrcid.ca/downloads/
Sust_Decision_Mak_Measur_Fed_Level_Sadler.ppt+Social+economic+environment+method+OR+too

l+OR+model+integration+OR+integrate+OR+integrative+%E2%80%9Csustainability+appraisal%E2%80%9D&cd=42&hl=en&ct=clnk&gl=nz

SchoBer, B., Helming, K. & Wiggering, H. 2010, "Assessing land use change impacts – a comparison of the SENSOR land use function approach with other frameworks" in Journal of Land Use Science, vol. 5, issue 2, pp. 159-178
HAVE ABSTRACT
Land use changes play a key role in sustainable, rural development. In decision-making for land use, sustainability impact assessment is therefore essential and must provide condensed key information about impact pathway relationships based on complex scientific analysis. For this purpose a number of science-based valuation frameworks exist, including ecosystem services as applied in the Millennium Ecosystem Assessment, landscape functions identified through landscape ecology and land use functions, a multifunctionality-based approach developed in the EU Integrated Project SENSOR (Sustainability Impact Assessment: Tools for environmental, social and economic effects of multifunctional land use in European regions). In this article, the three approaches are comparatively reviewed for their suitability for sustainability impact assessment of land use changes, using the following criteria: premises and perspectives, application in the science-policy interface, spatial and temporal references, consideration of the three sustainability dimensions and the role of land use. Core findings are that ecosystem services were biased towards the environmental dimension of sustainability and best suited for long-term projections. Landscape functions were aligned with the above sustainability concept and met prospective planning purposes. Land use functions were a pragmatic way for stakeholder-driven sustainability assessment of land use changes

Seghezzo, L. 2009, "The five dimensions of sustainability" in Environmental Politics, vol. 18, issue 4, pp. 539-556
HAVE FULL TEXT
Sustainability is usually seen as a guide for economic and social policymaking in equilibrium with ecological conditions. More than two decades after the World Commission on Environment and Development (WCED) defined 'sustainable development' and put the concept of sustainability on the global agenda, the concrete meaning of these terms and their suitability for specific cases remains disputed. A new conceptual framework to address sustainability issues is needed. The limitations of the WCED definition could be mitigated if sustainability is seen as the conceptual framework within which the territorial, temporal, and personal aspects of development can be openly discussed. Sustainability could be better understood in terms of 'Place', 'Permanence', and 'Persons'. Place contains the three dimensions of space, Permanence is the fourth dimension of time, and the Persons category represents a fifth, human dimension. The five-dimensional sustainability framework is arguably more inclusive, plural, and useful to outline specific policies towards sustainability

Sheate, W. 2010, 'The Evolving Nature of Environmental Assessment and Management: Linking Tools to Help Deliver Sustainability' in Tools, Techniques & Approaches for Sustainability. Collected Writings in Environmental Assessment Policy and Management, Sheate, W. (ed). World Scientific Publishing Co, Singapore.
BOOK – SOME TEXT ONLINE
http://books.google.co.nz/books?id=m3nCSDxWXUUC&pg=PA26&lpg=PA26&dq=Gibson,+R.+2006.+%E2%80%9CBeyond+the+Pillars:+Sustainability+Assessment+as+a+Framework+for+Effective+Integration+of&source=bl&ots=TZOk7DTD5h&sig=ttEN82YcjfbqdW91qAEyEGh1y2Y&hl=en&ei=b8kaTNLcIZGENr74-dwF&sa=X&oi=book_result&ct=result&resnum=8&ved=0CC4Q6AEwBw#v=onepage&q=Gibson%2C%20R.%202006.%20%E2%80%9CBeyond%20the%20Pillars%3A%20Sustainability%20Assessment%20as%20a%20Framework%20for%20Effective%20Integration%20of&f=false

Sheate, W. & Partidario, M. 2010, "Strategic approaches and assessment techniques – Potential for knowledge brokerage towards sustainability" in Environmental Impact Review, vol. 30, pp. 278-288
HAVE FULL TEXT
The role of science in policy and decision-making has been an issue of intensive debate over the past decade. The concept of knowledge brokerage has been developing in this context contemplating issues of communication, interaction, sharing of knowledge, contribution to common understandings, as well as to effective and efficient action. For environmental and sustainability policy and decision-making the discussion has addressed more the essence of the issue rather than the techniques that

can be used to enable knowledge brokerage. This paper aims to contribute to covering this apparent gap in current discussion by selecting and examining empirical cases from Portugal and the United Kingdom that can help to explore how certain environmental and sustainability assessment approaches can contribute, if well applied, to strengthen the science–policy link. The cases show that strategic assessment approaches and techniques have the potential to promote knowledge brokerage, but a conscious effort will be required to design in genuine opportunities to facilitate knowledge exchange and transfer as part of assessment processes

Thabrew, L., Wiek, A. & Ries, R. 2009, "Environmental decision making in multi-stakeholder contexts: applicability of life cycle thinking in development planning and implementation" in Journal of Cleaner Production, vol. 17, issue 1, pp. 67-76
HAVE FULL TEXT
According to the targets set for sustainability, integrating the principles of sustainable development into country policies and programs is one of the main goals for development projects. A major challenge in the development field is cross-sectoral integrated planning and achieving multi-stakeholder consensus for collaborative joint projects, especially when sustainability is a goal. This increases the complexity of the multi-stakeholder interaction in decision making and requires enhanced mechanisms for stakeholder participation, coordination, and commitment beyond narrow self-interest. A critical aspect in the decision making process is to enable stakeholders to not only interpret and make decisions based on expert judgments, but also to appropriately involve the relevant parties in the research and decision making process. Therefore, scientific analyses in multi-stakeholder contexts have to be more transparent, participatory, and stakeholder-based in order to provide useful information to assist responsible decision making. This paper presents an outline of a stakeholder-based life cycle assessment approach that can be used to support sustainable decision making in multi-stakeholder contexts. The framework is discussed and compared to other common methods used to support environmental decision making in development projects. We argue that the fundamental concept of life cycle thinking can be effectively used to incorporate stakeholders in the research and decision making process, which can lead to more comprehensive, yet achievable assessments in collaboration with stakeholders. Life cycle thinking is not just a way to examine environmental impacts of activities, but also a way to comprehend and visualize a broader set of upstream and downstream consequences of decisions in development planning and implementation. A life cycle framework including the mapping of stakeholder involvement at each activity in upstream and downstream stages would give stakeholders a holistic view that they otherwise may not have.

Thomson, C. et al. 2007, "The Role of Knowledge Management in Urban Sustainability Assessment", in M. Horner, C. Hardcastle, A. Price, & J. Bebbington (eds), International Conference on Whole Life Urban Sustainability and its Assessment, Glasgow
HAVE FULL TEXT
Urban sustainability assessment is increasingly being seen not just as a technical process to determine the likely sustainability performance of build projects but as a valuable tool in the mediation between the many associated stakeholders with their differing visions, numerous requirements and variations in expertise. This emerging role presents new and considerable challenges for the management of knowledge – during its generation, flow and capture - to ensure the meaningful engagement of such stakeholders in the decision-making process. This paper explores the role of knowledge management in aiding the delivery of urban sustainability assessment within development projects with multiple stakeholders. Outlined is an approach taken by the SUE- MoT research consortium in the development of a knowledge management system incorporated to supplement the practical application of a planned integrated sustainability assessment toolkit (ISAT). The system aims to deliver a knowledge support system (codification strategy) provided through an ICT resource for capturing, storing and transferring explicit and tacit knowledge generated during assessment. This will be integrated with a personalisation strategy developed to promote the necessary discourse between stakeholders deemed necessary to facilitate the transfer of knowledge required for assessment to function as a tool for mediation and 'social' learning. The integration of these strategies forms the basis for a wider knowledge management system that will be supplemented by a framework for managing sustainability assessment. A series of knowledge maps are to be developed to provide the foundations for understanding the contextual nature of the knowledge flow between stakeholders during assessment. This forms the basis for the development of the mechanisms provided through the system for stimulating knowledge transfer and to ensure the effective mediation of their varying

priorities. The paper argues that only by engaging with the contextual nature of this flow, can effective stakeholder engagement during the decisionmaking process become achievable

Thomson, C., El-Haram, M. & Hardcastle, C. 2009, "Managing knowledge of urban sustainability assessment" in Proceedings of the ICE – Engineering Sustainability, vol. 162, issue, 1, pp. 35-43
HAVE ABSTRACT
Urban sustainability assessment is increasingly being seen not just as a technical process to determine the likely sustainability performance of development projects but also as a valuable tool in mediation between many associated stakeholders with differing visions, numerous requirements and variations in expertise. This emerging role presents new and considerable challenges for the management of knowledge—during its generation, flow and capture—to ensure the meaningful engagement of stakeholders in the decision-making process. This paper explores the role of knowledge management in aiding the delivery of urban sustainability assessments within development projects with multiple stakeholders. The paper outlines the requirements identified by the sustainable urban environments—metrics, models and toolkits (SUEMOT) research consortium for developing a knowledge management system (KMS) designed to aid the management of knowledge generated during an assessment, through facilitation of its flow and transfer between those involved in an assessment and those involved in future assessments. The system is developed around integration of two knowledge management strategies—personalisation and codification. A methodology for the development of the system is presented, stressing the need to integrate strategies around a practical understanding of the role of a sustainability assessment within the development project lifecycle and the knowledge requirements associated with its context

Ticehurst, J. 2007, "A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales", in Environmental Modelling and Software, vol. 22, issue 8, pp. 1129-1139
HAVE FULL TEXT
Coastal lakes are ecosystems of significant value generating many ecological, social and economic benefits. Increasing demands resulting from urban development and other human activities within coastal lake catchments have the potential to result in their degradation and can lead to conflicts, for example between lake users and upstream communities. There are many techniques that can be used to integrate the variables involved in such conflicts including system dynamics, meta-modelling, and coupled component models, but many of these techniques are too complex for catchment managers to employ on a routine basis. The overall result is the potential to compromise the sustainability of these important ecosystems. This paper describes research to address this problem. It presents the development of an integrated model framework based on a Bayesian network (Bn). Bns are used to assess the sustainability of eight coastal lake-catchment systems, located on the coast of New South Wales (NSW), Australia. The paper describes the potential advantages in the use of Bns and the methods used to develop their frameworks. A case study application for the Cudgen Lake of northern NSW is presented to illustrate the techniques. The case study includes a description of the relevant management issues being considered, the model framework and the techniques used to derive input data. Results for the case study application and their implications for management are presented and discussed. Finally, the directions for future research and a discussion of the applicability of Bn techniques to support management in similar situations are proffered.

Tuinstra, W., Jäger, J. & Weaver, P. 2008, "Learning and evaluation in Integrated Sustainability Assessment" in International Journal of Innovation and Sustainable Development, Vol. 3, Nos. 1–2, pp.128–152.
HAVE FULL TEXT
This paper explains why learning, especially social learning, and evaluation are essential and integral parts of Integrated Sustainability Assessment (ISA). This is linked to the intent and purpose of ISA as a process that seeks to achieve transformative outcomes, including through issue reframing. The goals of ISA require a participatory process, adaptability and an iterative architecture with, at the end of each ISA cycle, a phase of purposefully structured evaluation of learning achieved. The paper explains how the evaluation of learning outcomes has been structured in the MATISSE project and reports findings from the first round of evaluation

Turnpenny, J. et al. 2008, "Why is integrating policy assessment so hard? A comparative analysis of the institutional capacities and constraints" in Journal of Environmental Planning and Management, vol. 51, issue 6, pp. 759-775

HAVE FULL TEXT
Widely advocated as a means to make policy making more integrated, policy assessment remains weakly integrated in practice. But explanations for this shortfall, such as lack of staff training and resources, ignore more fundamental institutional factors. This paper identifies institutional capacities supporting and constraining attempts to make policy assessment more integrated. A comparative empirical analysis of functionally equivalent assessment systems in four European jurisdictions finds that there are wide-ranging institutional constraints upon integration. These include international policy commitments, the perception that assessment should support rather than determine policy, organisational traditions, and the sectorisation of policy making. This paper concludes by exploring the potential for altering these institutions to make policy assessment more integrated

Videira, N. et al. 2010, "A participatory modelling approach to support integrated sustainability assessment process" in Systems Research and Behavioral Science, vol. 27, issue 4, pp. 446-460
HAVE FULL TEXT
Several authors have described the usefulness of participatory system dynamics approaches in environmental decision-making processes, particularly in supporting problem scoping and policy analysis. This paper explores how these approaches may be expanded to provide a coherent, deliberative platform, which structures Integrated Sustainability Assessments (ISA) of policy proposals. The proposed ISA framework forms a five-stage feedback and learning process including scoping, visioning, model building, simulation/assessment and monitoring. We discuss these elements considering a set of best practice principles for participation in environmental assessment and decision making. Subsequently, we lay out a roadmap for implementing and testing the framework. We conclude that participatory modelling has a strong potential for supporting ISA processes. It allows stakeholders to build alternative policies, to reflect on their long-term dynamics, and to gain insights on the interrelationships underlying persistent sustainability problems

Waheed, B., Khan, F. & Veitch, B. 2009, "Linkage-based Frameworks for Sustainbaility Assessment: Making a Case for Driving-Force-Pressure-State-Exposure-Effect-Action (DPSEEA) Frameworks", in Sustainability, vol. 1, pp. 441-463
HAVE FULL TEXT
"The main objective of this paper is to discuss different approaches, identify challenges, and to select a framework for delivering effective sustainability assessments. Sustainable development is an idealistic concept and its assessment has always been a challenge. Several approaches, methodologies and conceptual frameworks have been developed in various disciplines, ranging from engineering to business and to policy making. The paper focuses mainly on various linkage-based frameworks and demonstrates that the driving force-state-exposure-effect-action (DPSEEA) framework can be used to achieve sustained health benefits and environmental protection in accordance with the principles of sustainable development, especially because of its resemblance to the environmental risk assessment and management paradigms. The comparison of linkage-based frameworks is demonstrated through an example of sustainability in a higher educational institution."

Weaver, P. & Rotmans, J. 2006, "Integrated sustainability assessment: What, why and how" in International Journal of Innovation and Sustainable Development, Vol. 1, No. 4, pp.284–303.
HAVE ABSTRACT, CAN'T FIND FULL TEXT, asked Cawthron
This paper describes a new conceptualisation of sustainability assessment as an integrative and active process at the science-policy-society interface and its implementation as exemplified in case studies within the Methods and Tools for Integrated Sustainability Assessment (MATISSE) project. Integrated Sustainability Assessment (ISA) is defined within the MATISSE project as a cyclical, participatory process of scoping, envisioning, experimenting and learning through which a shared interpretation of sustainability for a specific context is developed and applied in an integrated manner in order to explore solutions to persistent problems. ISA is strategic, sustainability-oriented, constructive and potentially transformative. Its key role is to explore the opportunity-creation and problem-solving potential of framing contexts other than those in place, such as alternative institutions, technologies, spatial and temporal arrangements, price relations and associated policy regimes.

Weaver, P. & Jordan, A. 2008, "What roles are there for sustainability assessment in the policy process?" in International Journal of Innovation and Sustainable Development, Vol. 3, Nos. 1–2, pp.9–32.
HAVE FULL TEXT

This paper reviews some of the attempts that have been made in the EU to include sustainability concerns in ex ante policy assessment processes and explores why these are generally deemed to be ineffective. It argues that the dominant policy frames are not oriented towards sustainable development. Traditional policy assessment procedures, which are designed and used to screen policy proposals, are unable to install sustainability considerations into proposals retroactively. A 'gap' in the toolbox of assessment methods, therefore, exists for a form of sustainability assessment able to give a strategic sustainability orientation to policy-making. Within the MATISSE project, this form of assessment is termed 'Integrated Sustainability Assessment' (ISA). This paper describes ISA and discusses where and when in the policy process it could be used most effectively

Weidema, B. 2006, "The Integration of Economic and Social Aspects in Life Cycle Assessment", in The International Journal of Life Cycle Assessment, vol. 11, no. 1, pp. 89-96
HAVE FULL TEXT
Goal, Scope and Background   Although both cost-benefit analysis (CBA) and life cycle assessment (LCA) have developed from engineering practice, and have the same objective of a holistic ex-ante assessment of human activities, the techniques have until recently developed in relative isolation. This has resulted in a situation where much can be gained from an integration of the strong aspects of each technique. Such integration is now being prompted by the more widespread use of both CBA and LCA on the global arena, where also the issues of social responsibility are now in focus. Increasing availability of data on both biophysical and social impacts now allow the development of a truly holistic, quantitative environmental assessment technique that integrates economic, biophysical and social impact pathways in a structured and consistent way. The concept of impact pathways, linking biophysical and economic inventory results via midpoint impact indicators to final damage indicators, is well described in the LCA and CBA literature. Therefore, this paper places specific emphasis on how social aspects can be integrated in LCA. Methods and Results. With a starting point in the conceptual structure and approach of life cycle impact assessment (LCIA), as developed by Helias Udo de Haes and the SETAC/UNEP Life Cycle Initiative, the paper identifies six damage categories under the general heading of human life and well-being. The paper proposes a comprehensive set of indicators, with units of measurement, and a first estimate of global normalisation values, based on incidence or prevalence data from statistical sources and severity scores from health state analogues. Examples are provided of impact chains linking social inventory indicators to impacts on both human well-being and productivity. Recommendation and Perspective It is suggested that human well-being measured in QALYs (Quality Adjusted Life Years) may provide an attractive single-score alternative to direct monetarisation.

Whitmarsh, L., Swartling, A. & Jager, J. 2009, "Participaton of Experts and Non-Experts in a Sustainability Assessment of Mobility" in Environmental Policy and Governance, vol. 19, pp. 232-250
HAVE FULL TEXT
The complexity, ambiguity and subjectivity that surround persistent problems of unsustainability, such as mobility, highlight the importance of stakeholder engagement in both knowledge production and policy development. This paper reports on research within the EU-funded MATISSE project to develop tools and methods for Integrated Sustainability Assessment (ISA), a novel interdisciplinary and participatory approach to sustainability strategy development. Two different methods – expert focus groups and citizen deliberative workshops – were employed to elicit knowledge and preferences of European stakeholders in respect of sustainable mobility. Findings from these exercises indicate areas of both convergence and divergence in the visions of sustainable mobility futures depicted by different stakeholder groups. Stakeholders agreed on the need to address problems of unsustainability in the transport sector, and identified broadly similar environmental, social and economic criteria for sustainable transport. Amenity of transport was more important for citizens, while experts focussed on pragmatic and technological issues. Both groups favoured modal shift and novel technologies, and citizens also supported demand reduction measures and choices; however, a range of barriers to achieving sustainable mobility was also identified by participants. Stakeholder feedback suggests the process was valuable and acted as a forum for social learning and the co-production of knowledge by citizens and experts, while at the same time empowering these groups to participate in an important social issue such as transport. The value and limitations of these methods for ISA are discussed and avenues for further research proposed.

Yigitcanlar, T. & Dur, F. 2010, "Developing a Sustainability Assessment Model: The Sustainable Infrastructure, Land-Use, Environment and Transport Model" in Sustainability, vol. 2, issue 1. pp, 321-340
HAVE FULL TEXT
Measuring the comparative sustainability levels of cities, regions, institutions and projects is an essential procedure in creating sustainable urban futures. This paper introduces a new urban sustainability assessment model: "The Sustainable Infrastructure, Land-use, Environment and Transport Model (SILENT)". The SILENT Model is an advanced geographic information system and indicator-based comparative urban sustainability indexing model. The model aims to assist planners and policy makers in their daily tasks in sustainable urban planning and development by providing an integrated sustainability assessment framework. The paper gives an overview of the conceptual framework and components of the model and discusses the theoretical constructs, methodological procedures, and future development of this promising urban sustainability assessment model.

Part 2: Literature search on integrated sustainability indicator systems for urban development scenario assessment in the context of receiving water bodies for urban storm water

Armitage, D. et al. 2009, "Adaptive co-management for social-ecological complexity", in Frontiers in Ecology and the Environment, vol. 7, no. 2, pp. 95-102
HAVE FULL TEXT
Building trust through collaboration, institutional development, and social learning enhances efforts to foster ecosystem management and resolve multi-scale society–environment dilemmas. One emerging approach aimed at addressing these dilemmas is adaptive co-management. This method draws explicit attention to the learning (experiential and experimental) and collaboration (vertical and horizontal) functions necessary to improve our understanding of, and ability to respond to, complex social–ecological systems. Here, we identify and outline the core features of adaptive co-management, which include innovative institutional arrangements and incentives across spatiotemporal scales and levels, learning through complexity and change, monitoring and assessment of interventions, the role of power, and opportunities to link science with policy

Bagheri, A. & Hjorth, P. 2006, "Planning for sustainable development: a paradigm shift towards a process-based approach" in Sustainable Development, vol. 15, issue 2, pp. 83-96
HAVE FULL TEXT
Prevailing approaches of planning and strategy making, which traditionally deal with the states of systems in terms of fixed goals, fail to acknowledge the process nature of sustainable development. Using a system dynamics approach and relying on the concept of viability loops, the paper aims to illustrate a practical implementation of sustainable development with an urban water system as an example. It argues that planning for sustainable development should be 'process-based' – rather than 'fixedgoal' – oriented. Unlike the traditional approaches of strategy making to set fixed goals related to either supply-side and/or demand-side management, it is argued that triggering a social learning process with full involvement of all stakeholders and planners in the process would be the most suitable strategy for sustainable development. To this end, backcasting is recommended as a suitable tool and the process of model building is regarded as a means of learning rather than forecasting

Becker, J. 2010, "Use of backcasting to integrate indicators with principles of sustainability" in International Journal of Sustainable Development and World Ecology, vol. 17, issue 3, pp. 189-197
HAVE ABSTRACT
There are innumerable indicators that can be used to evaluate sustainable development and these can serve to inform and guide progress. However, they are often selected without considering their effect on each other and the achievement of sustainability. Indeed, two decades after the definition of sustainable development, there is still little agreement on principles of sustainability, nor the indicators to represent them. Part of the difficulty lies in working from a present unsustainable state to an uncertain future horizon. An alternative is to apply backcasting, which starts from a desired outcome and develops scenarios for achieving it. This paper presents a framework of sustainability principles to help operationalize indicators and increase their relevance. It outlines how this can be done using backcasting to evaluate existing indicators or to select those that relate positively to the framework principles. Although there is no formula for achieving sustainability, this process provides a format for

discussion in selecting indicators and in ensuring their continued relevance to sustainable development progress

Bell, S. & Morse, S. 2008, Sustainability Indicators: Measuring the immeasurable? Earthscan, London, Sterling, VA).
CAN'T FIND

Bennett, E. et al. 2005, "A Systems Model Approach to Determining Resilience Surrogates for Case Studies", in Ecosystems, vol. 8, no. 8, pp. 945-957
HAVE FULL TEXT
Resilience theory offers a framework for under standing the dynamics of complex systems. How ever, operationalizing resilience theory to develop and test empirical hypotheses can be difficult. We present a method in which simple systems models are used as a framework to identify resilience surrogates for case studies. The process of constructing a systems model for a particular case offers a path for identifying important variables related to sys tem resilience, including the slowly-changing variables and thresholds that often are keys to understanding the resilience of a system. We develop a four-step process for identifying resilience Surrogates through development of systems models. Because systems model development is often a difficult step, we summarize four basic existing systems models and give examples of how each may be used to identify resilience surrogates. The construction and analysis of simple systems models provides a useful basis for guiding and directing the selection of surrogate variables that will offer appropriate empirical measures of resilience.

Berkes, F. & Seixas, C. 2005, "Building Resilience in Lagoon Social-Ecological Systems: A Local-Level Perspective", in Ecosystems, vol. 8, no. 8, pp. 967-974
HAVE FULL TEXT
Building resilience in integrated human and nature systems or social–ecological systems (SES) is key for sustainability. Therefore, developing ways of assessing resilience is of practical as well as theoretical significance. We approached the issue by focusing on the local level and using five lagoon systems from various parts of the world for illustration. We used a framework based on four categories of factors for building resilience: (1) learning to live with change and uncertainty; (2) nurturing diversity for reorganization and renewal; (3) combining different kinds of knowledge; and (4) creating opportunity for self-organization. Under each category, the cases generated a number of items for building resilience, and potential surrogates of resilience, that is, variables through which the persistence of SES emerging through change can be assessed. The following factors were robust across all five lagoon SES cases: learning from crisis, responding to change, nurturing ecological memory, monitoring the environment, and building capacity for self-organization and conflict management.

Binder, C., Feola, G. & Steinberger, J. 2010, "Considering the normative, systematic, procedural dimensions in indicator-based sustainability assessments in agriculture" in Environmental Impact Assessment Review, vol. 30, issue 2, pp. 71-81
HAVE FULL TEXT
This paper develops a framework for evaluating sustainability assessment methods by separately analyzing their normative, systemic and procedural dimensions as suggested by Wiek and Binder [Wiek, A, Binder, C. Solution spaces for decision-making - a sustainability assessment tool for city-regions. The framework is then used to characterize indicator-based sustainability assessment methods in agriculture. For a long time, sustainabillity assessment in agriculture has focused mostly on environmental and technical issues, thus neglecting the economic and, above all, the social aspects of sustainability. the multifunctionality of agriculture and the applicability of the results. In response to these shortcomings, several integrative sustainability assessment methods have been developed for the agricultural sector. This paper reviews seven of these that represent the diversity of tools developed in this area. The reviewed assessment methods can be categorized into three types: (i) top-down farm assessment methods; (ii) top-down regional assessment methods with some stakeholder participation; (iii) bottom-up, integrated participatory or transdisciplinary methods with stakeholder participation throughout the process. The results readily show the trade-offs encountered when selecting an assessment method. A clear, standardized, top-down procedure allows for potentially benchmarking and comparing results across regions and sites. However, this comes at the cost of system specificity. As the top-down methods often have low stakeholder involvement, the application and implementation of the results might be difficult. Our analysis suggests that to include

the aspects mentioned above in agricultural sustainability assessment, the bottom-up, integrated participatory or transdisciplinary methods are the most suitable ones

Boggia, A. & Cortina, C. 2010, “Measuring sustainable development using a multi-criteria model: A case study” in Journal of Environmental Management, In Press
HAVE FULL TEXT
This paper shows how Multi-criteria Decision Analysis (MCDA) can help in a complex process such as the assessment of the level of sustainability of a certain area. The paper presents the results of a study in which a model for measuring sustainability was implemented to better aid public policy decisions regarding sustainability. In order to assess sustainability in specific areas, a methodological approach based on multi-criteria analysis has been developed. The aim is to rank areas in order to understand the specific technical and/or financial support that they need to develop sustainable growth. The case study presented is an assessment of the level of sustainability in different areas of an Italian Region using the MCDA approach. Our results show that MCDA is a proper approach for sustainability assessment. The results are easy to understand and the evaluation path is clear and transparent. This is what decision makers need for having support to their decisions. The multi-criteria model for evaluation has been developed respecting the sustainable development economic theory, so that final results can have a clear meaning in terms of sustainability

Carpenter, S. et al. 2005, “Surrogates for Resilience of Social-Ecological Systems”, in Ecosystems, vol. 8, no. 8, pp. 941-944
NO ABSTRACT, EXCERPT FROM INTRODUCTION…
From its roots in ecology, resilience (Holling 1973) has more recently been applied to social-ecological systems, or SES. Theories of changing resilience explicitly address the persistence or breakdown of diverse states of complex systems (Gunderson and Holling 2002). These ideas have attracted interest from interdisciplinary research groups interested in change, conservation or restoration of SES (Berkes and others 2003; Scheffer and others 2003). From a practical standpoint, resilience theory provides a conceptual foundation for sustainable development (Folke and others 2002). The transition from theory to practice, however, requires assessment or estimation of resilience (Carpenter and others 2001). So far, there is little experience with estimating resilience of SES, and little understanding of the sensitivity of resilience measures to changes in SES. This shortage of practical field experience is a barrier to building understanding through empirical study of resilience in SES. Direct measurement of resilience is difficult be cause it requires measuring the thresholds or boundaries that separate alternate domains of dynamics for SES. The only sure way to detect a threshold in a complex system is to cross it (Carpenter 2003). Yet threshold-crossings do not occur very often. In the natural sciences, much understanding of thresholds has come from deliberate manipulations of ecosystems, or before/after studies of large disturbances (Turner and Dale Groffman and others 2005). In interdisciplinary science, it may be impossible, unethical or both to induce a threshold-crossing in an SES. It may be possible to assess SES thresholds retrospectively, through historical analysis of case studies. But, especially in applications of resilience to pressing environmental problems, we may be interested in present-day resilience or future resilience of an SES. Indirect inferences are unavoidable in such circumstances. Resilience measures differ from traditional eco logical indicators (for example, National Research Council 2000) in several respects. For this reason, we use the word "surrogates" instead of "indicators". By referring to surrogates, we acknowledge that important aspects of resilience in SES may not be directly observable, but must be inferred indirectly. We also acknowledge that resilience and its relationship to the surrogates may change over time. Resilience surrogates should correspond in a specified way to theoretical aspects of resilience. According to resilience theory, the aspects of a system that confer resilience depend on context (Holling 2001). These change: over time, with spatial configuration, among the diverse ecosystems that may comprise a regional SES, and among groups of people participating in the SES. Consequently, the relationship between resilience and any particular surrogate may be dynamic, complex, and multidimensional. In general, practitioners will need a suite of resilience surrogates that jointly represent the key features of resilience…

Clark, J. & Clarke, R. 2010, “Local sustainability initiatives in English National Parks: What role for adaptive governance?” in Land Use Policy, In Press
HAVE FULL TEXT
Adaptive governance has assumed growing importance in natural resource management literatures, emphasising learning and adaptation among actors at different political administrative levels and

geographic scales as a precondition for the emergence of sustainable development. Here we assess this claim by examining five case studies of ‘good practice’ in sustainability, drawn from a national survey conducted in English National Parks. Specifically, we evaluate whether (1) adaptive governance characteristics are present in these ‘good practice’ initiatives, and (2) what governance role, if any, National Park Authorities have played in mediating individual and collective activities and behaviours within these projects at different levels and scales. We conclude with a critical assessment of the capacity of the adaptive governance approach to furnish new understandings of sustainable development initiatives in English NPs

Cobiac, L. 2006, “Selection of Indicators for sustainable management of an urban stormwater system”, 2nd IWA Leading Edge Conference on Sustainability, IWA Publishing, London
BOOK – SOME PAGES AVAILABLE ONLINE, COUNDN’T COPY ABSTRACT, LOOKS INTERESTING…James wants; can’t find
http://books.google.co.nz/books?hl=en&lr=&id=ajA4ZN-KIvYC&oi=fnd&pg=PA25&dq=method+integrate+OR+integration+OR+integrative+%22sustainability+assessment%22+-climate+-medical&ots=xseo3iIdrM&sig=UE7UKIKF9xq1yGc1gwIAbwnM5Sk#v=onepage&q=method%20integrate%20OR%20integration%20OR%20integrative%20%22sustainability%20assessment%22%20-climate%20-medical&f=false

Cumming, G. et al. 2005, “An Exploratory Framework for the Empirical Measurement of Resilience”, in Ecosystems, vol. 8, no. 8, pp. 975-987
HAVE FULL TEXT
Deliberate progress towards the goal of long-term sustainability depends on understanding the dynamics of linked social and ecological systems. The concept of social-ecological resilience holds promise for interdisciplinary syntheses. Resilience is a multifaceted concept that as yet has not been directly operationalized, particularly in systems for which our ignorance is such that de tailed, parameter-rich simulation models are difficult to develop. We present an exploratory framework as a step towards the operationalization of resilience for empirical studies. We equate resilience with the ability of a system to maintain its identity, where system identity is defined as a property of key components and relationships (networks) and their continuity through space and time. Innovation and memory are also fundamental to understanding identity and resilience. By parsing our systems into the elements that we subjectively consider essential to identity, we obtain a small set of specific focal variables that reflect changes in identity. By assessing the potential for changes in identity under specified drivers and perturbations, in combination with a scenario-based approach to considering alternative futures, we obtain a surrogate measure of the current resilience of our study system as the likelihood of a change in system identity under clearly specified conditions, assumptions, drivers and perturbations. Although the details of individual case studies differ, the concept of identity provides a level of generality that can be used to compare measure of resilience across cases. Our approach will also yield insights into the mechanisms of change and the potential consequences of different policy and management decisions, providing a level of decision support for each case study area

Cundill, G. & Fabricius, C. 2010, “Monitoring the Governance Dimension of Natural Resource Co-Management” in Ecology and Society, vol. 15, issue 1
HAVE FULL TEXT
The governance outcomes of natural resource co-management have been neither systematically monitored nor rigorously assessed. We identified system attributes and key variables that could form the basis for monitoring the governance dimension of adaptive co-management. A methodology for collaboratively monitoring these system attributes and key variables was tested in four localities in South Africa. Our results suggest that creating the conditions that facilitate self-organization, and particularly cross-scale institutional linkages, is the major challenge facing attempts to initiate adaptive co-management. Factors requiring greater attention include community perceptions of support from outside agencies, access to long-term funding for adaptive decision making, and access to reliable information about changes in natural resources and legal options for the formation of decision-making bodies. Longterm and well-funded social facilitation is key to achieving this
Diedrich, A., Tintore, J. & Navines, F. 2010, “Balancing science and society through establishing indicators for integrated coastal zone management in the Balearic Islands” in Marine Policy, vol. 34, no. 4, pp. 772-781
HAVE FULL TEXT

This paper explores the process by which indicators may be developed as tools for communicating science to decision-makers using the participatory approach demonstrated by the Balearic Indicators Project. This initiative reflects a series of compromises considered necessary to achieve the objective of generating an indicator system that is scientifically viable, comparative internationally yet locally relevant, and to facilitate its implementation. The article highlights questions regarding the utility of science for addressing current global issues related to sustainability and why science often fails to promote change at the societal level

Dizdaroglu, D. et al 2009, “An Indexing Model for Stormwater Quality Assessment: Sustainable Stormwater Management in the Gold Coast”, Proceedings of the 2nd International Urban Design Conference Survival: Implementing Tomorrow's City, 2-4 September 2009, Gold Coast, Queensland
HAVE FULL TEXT
In the age of climate change and rapid urbanisation, stormwater management and water sensitive urban design have become important issues for urban policy makers.
This paper reports the initial findings of a research study that develops an indexing model for assessing stormwater quality in the Gold Coast

Duit, A. et al. 2010, “Governance, complexity, and resilience” in Global Environmental Change, vol. 20, issue 3, pp. 363-368
HAVE FULL TEXT
This special issue brings together prominent scholars to explore novel multilevel governance challenges posed by the behavior of dynamic and complex social-ecological systems. Here we expand and investigate the emerging notion of “resilience” as a perspective for understanding how societies can cope with, and develop from, disturbances and change. As the contributions to the special issue illustrate, resilience thinking in its current form contains substantial normative and conceptual difficulties for the analysis of social systems. However, a resilience approach to governance issues also shows a great deal of promise as it enables a more refined understanding of the dynamics of rapid, interlinked and multiscale change. This potential should not be underestimated as institutions and decision-makers try to deal with converging trends of global interconnectedness and increasing pressure on social-ecological systems

Engle, N. & Lemos, M. “Governance determinants and indicators for resilience and adaptive capacity in Brazilian river basin management”
[http://www.openmeeting2009.org/pdf_files/Pdf%20papers/Microsoft%20Word%20-%20IHDP2009_Full_Paper_Engle_and_Lemos_final.pdf]
HAVE FULL TEXT
Recent developments within the adaptation field have revealed the importance of institutions and governance mechanisms for building both adaptive capacity and resilience. Scholarship suggests that adaptive governance is a pivotal approach for increasing a system’s flexibility and robustness in the face of climate change. This paper assesses adaptive governance in Brazilian river basins that have experienced recent decentralization of water management. Assessments are based on a survey of 18 river basin committees throughout Brazil. We create a governance index to cluster the basins into three categories across seven governance and institutional indicator variables. The results are then compared to interview data within two of the basins to examine the governance model’s influence on adaptive capacity and resilience. We find that there are considerable differences between the clustering of basins’ governance approaches. The in-depth interviews in the two case study basins suggest higher adaptive capacity and resilience in the basin with more adaptive governance implementation. Perhaps most importantly, there is evidence indicating that some of mechanisms in the adaptive governance model might actually vary in opposite directions. This finding is particularly evident in the ‘equality of knowledge and information use’ variable. That is, it appears that in the decentralization of Brazilian river basin management, equality of knowledge and information is not associated with high levels other adaptive governance variables. In order to offer insight for how to reverse this trend, we conclude with a preliminary exploration into what factors lead to knowledge and information use

Engle, N. & Lemos, M. 2010, “Unpacking governance: Building adaptive capacity to climate change of river basins in Brazil” in Global Environmental Change-Human and Policy Dimensions, vol. 20, issue 1, pp. 4-13
HAVE FULL TEXT

Governance and institutions are critical determinants of adaptive capacity and resilience. Yet the make-up and relationships between governance components and mechanisms that may or may not contribute to adaptive capacity remain relatively unexplored empirically. This paper builds on previous research focusing on integrated water resources management in Brazil to 'unpack' water governance mechanisms that may shape the adaptive capacity of water systems to climatic change. We construct a river basin index to characterize governance approaches in 18 Brazilian river basins, apply a reliability test to assess the validity of these governance indicators, and use in-depth qualitative data collected in a subsample of the basins to explore the relationship between the governance indicators and adaptive capacity. The analysis suggests a positive relationship between integrated water governance mechanisms and adaptive capacity. In addition, we carry out a cluster analysis to group the basins into types of governance approaches and further unveil potential relationships between the governance variables and overall adaptive capacities. The cluster analysis indicates that tensions and tradeoffs may exist between some of the variables, especially with equality of decision making and knowledge availability; a finding that has implications for decision makers aiming to build adaptive capacity and resilience through governance and institutional means

Fabricius, C. et al. 2007, “Powerless spectators, coping actors, and adaptive co-managers: a synthesis of role of communities in ecosystem management” in Ecology and Society, vol. 12, issue 1
HAVE FULL TEXT
We provide a synthesis of the papers in the Special Issue, the Communities Ecosystems and Livelihoods component of the Millennium Ecosystem Assessment (MA), and other recent publications on the adaptive capacity of communities and their role in ecosystem management. Communities adapt because they face enormous challenges due to policies, conflicts, demographic factors, ecological change, and changes in their livelihood options, but the appropriateness of their responses varies. Based on our synthesis, three broad categories of adaptive communities are identified. "Powerless spectator" communities have a low adaptive capacity and weak capacity to govern, do not have financial or technological options, and lack natural resources, skills, institutions, and networks. "Coping actor" communities have the capacity to adapt, but are not managing social - ecological systems. They lack the capacity for governance because of lack of leadership, of vision, and of motivation, and their responses are typically short term. "Adaptive manager" communities have both adaptive capacity and governance capacity to sustain and internalize this adaptation. They invest in the long-term management of ecosystem services. Such communities are not only aware of the threats, but also take appropriate action for long-term sustainability. Adaptive co-management becomes possible through leadership and vision, the formation of knowledge networks, the existence or development of polycentric institutions, the establishment and maintenance of links between culture and management, the existence of enabling policies, and high levels of motivation in all role players. Adaptive co-managers are empowered, but empowerment is a consequence of the capacity for governance and the capacity to adapt, rather than a starting point. Communities that are able to enhance their adaptive capacity can deal with challenges such as conflicts, make difficult trade-offs between their short- and long-term well- being, and implement rules for ecosystem management. This improves the capacity of the ecosystem to continue providing services

Fischer, J. 2009, “Integrating resilience thinking and optimisation for conservation” in Trends in Ecology and Evolution, vol. 24, issue 10, pp. 549-554
HAVE FULL TEXT
Conservation strategies need to be both effective and efficient to be successful. To this end, two bodies of research should be integrated, namely ‘resilience thinking’ and ‘optimisation for conservation,’ both of which are highly policy relevant but to date have evolved largely separately. Resilience thinking provides an integrated perspective for analysis, emphasising the potential of nonlinear changes and the interdependency of social and ecological systems. By contrast, optimisation for conservation is an outcome-oriented tool that recognises resource scarcity and the need to make rational and transparent decisions. Here we propose that actively embedding optimisation analyses within a resilience-thinking framework could draw on the complementary strengths of the two bodies of work, thereby promoting cost-effective and enduring conservation outcomes

Foley, B. & Daniell, T. 2004, “Sustainability tool for intrasectoral and intersectoral water resources decision making”, in Australian Journal of Water Resources. Vol. 8, pp. 11-19
HAVE ABSTRACT James wants; can’t find – asking Cawthron to get

[NOTE authors from School of Civil & Environmental Engineering, The University of Adelaide SA bfoley@civeng.adelaide.edu.au trevord@civeng.adelaide.edu.au]
The concept of and need for sustainable water resource systems is widely accepted. Organisations at all levels both locally and globally and in all sectors from water resource management to transportation have attempted to define a set of criteria for sustainable systems. An examination of these criteria has revealed that there is a core set of sustainability goals that encapsulate the economic, social and environmental components of a sustainable system within any sector. However, the power of these goals can only be realised when they are used as a tool within a process which measures the effectiveness of different projects and schemes to meet the sustainability objectives, and upon which decisions are made.

Folke, C. et al. 2005, "Adaptive Governance of Social-Ecological Systems" in Annual of Environment and Resources, vol. 30, pp.441-473
HAVE FULL TEXT
We explore the social dimension that enables adaptive ecosystem-based management. The review concentrates on experiences of adaptive governance of social-ecological systems during periods of abrupt change (crisis) and investigates social sources of renewal and reorganization. Such governance connects individuals, organizations, agencies, and institutions at multiple organizational levels. Key persons provide leadership, trust, vision, meaning, and they help transform management organizations toward a learning environment. Adaptive governance systems often self-organize as social networks with teams and actor groups that draw on various knowledge systems and experiences for the development of a common understanding and policies. The emergence of "bridging organizations" seem to lower the costs of collaboration and conflict resolution, and enabling legislation and governmental policies can support self-organization while framing creativity for adaptive co-management efforts. A resilient social-ecological system may make use of crisis as an opportunity to transform into a more desired state

Folke, C. 2006, "Resilience: The emergence of a perspective for social-ecological systems analyses" in Global Environmental Change, vol. 16, issue 3, pp. 253-267
HAVE FULL TEXT
The resilience perspective is increasingly used as an approach for understanding the dynamics of social–ecological systems. This article presents the origin of the resilience perspective and provides an overview of its development to date. With roots in one branch of ecology and the discovery of multiple basins of attraction in ecosystems in the 1960–1970s, it inspired social and environmental scientists to challenge the dominant stable equilibrium view. The resilience approach emphasizes non-linear dynamics, thresholds, uncertainty and surprise, how periods of gradual change interplay with periods of rapid change and how such dynamics interact across temporal and spatial scales. The history was dominated by empirical observations of ecosystem dynamics interpreted in mathematical models, developing into the adaptive management approach for responding to ecosystem change. Serious attempts to integrate the social dimension is currently taking place in resilience work reflected in the large numbers of sciences involved in explorative studies and new discoveries of linked social–ecological systems. Recent advances include understanding of social processes like, social learning and social memory, mental models and knowledge–system integration, visioning and scenario building, leadership, agents and actor groups, social networks, institutional and organizational inertia and change, adaptive capacity, transformability and systems of adaptive governance that allow for management of essential ecosystem services

Folke, C. 2007, "Social-ecological systems and adaptive governance of the commons" in Ecological Research, vol. 22, issue 1, pp. 14-15
HAVE FULL TEXT

Gamper, C. & Turcanu, C. 2007, "On the governmental use of multi-criteria analysis" in Ecological Economics, vol. 62, issue 2, pp. 298-307
HAVE FULL TEXT
Public decision making, especially about our natural environment, is inherently exposed to a high conflict potential. The necessity to capture the complex context has led to an increasing request for decision analytic techniques as support for the decision process. Multi-criteria analysis (MCA) is deemed to overcome the shortcomings of traditional decision-support tools used in economics, such as cost-benefit (CBA) or cost-effectiveness analysis (CEA). This is due, among other, to its ability of

dealing with qualitative criteria (e.g. sensitive ecological factors), as well as with uncertainties about current or future impacts. Unlike CBA or CEA, MCA is rarely required by national laws or directives. Nonetheless, a number of recent MCA applications were supported by public authorities who either initiated or directly participated in such analyses. Given the theoretical assumptions about MCA's potential to support complex decision problems, as is often the case for environmental or sustainability policies, the key concern in our paper is to evaluate whether this potential has already been recognised in public decision making. For limitation purposes, the present work focuses on real-life case studies reported during the last decade with an insight in the initiation, the actors involved and the importance of the MCA results in the decision process. We argue that the significance and role played by MCA so far reaches beyond its current legal requirements

Gunningham N 2008. Innovative governance and regulatory design: managing water resources, Landcare Research Contract Report: LC0708/137, Prepared for FRST, Landcare Research New Zealand Limited.
HAVE FULL TEXT

Hayward, L. & De Wet, B. 2009, “Sustainability assessment: a tool for planning for sustainability as a desired outcome for a proposed development”, International Association for Impact Assessment (South Africa) Conference, Wilderness, South Africa, 23-26 August, 2009. pp 1-9
HAVE FULL TEXT
This paper presents a theoretical framework for planning for sustainability for any proposed development project. The objective of this framework is to foster and preserve the social ecological system in which the proposed development project is to occur so that the system remains dynamic, adaptive, resilient and durable through time. The overall approach to the framework is the understanding of the function and relationships within the social ecological system in which the proposed development is to occur. The system is analysed through stakeholder engagement and expert analysis. Sustainability is assessed through the development of sustainability principles and criteria with sustainability indicators to measure the progression. The framework has developed through a long history of impact assessment and strategic environmental assessment in the environmental sector.

Hemphill, L., Berry J. & McGreal, S. 2004, “An Indicator-based Approach to Measuring Sustainable Urban Regeneration Performance: Part 1, Conceptual Foundations and Methodological Framework”, in Urban Studies, vol. 41, No. 4. pp, 725-755
HAVE FULL TEXT
“This paper evaluates the use of indicators as a means of measuring the performance of regeneration against sustainability criteria. The merits of sustainability indicators are explored with a discussion of the indicator selection process and the derivation of a points scoring framework. The extent to which the sustainability performance of regeneration projects can be evaluated and subsequently benchmarked is examined. Conclusions are drawn on the robustness of the indicators selected, the versatility of the points scoring framework in capturing the sustainability performance of regeneration projects and the potential to identify 'best' practice. The application of the model and associated indicators is the subject of the subsequent paper in this journal issue.”

Hemphill, L., Berry J. & McGreal, S. 2004, “An Indicator-based Approach to Measuring Sustainable Urban Regeneration Performance: Part 2, Empirical Evaluation and Case-study Analysis”, in Urban Studies, vol. 41, No. 4. pp, 757-772
HAVE FULL TEXT
“This paper examines the sustainability of current urban regeneration practice, through the application of weighted indicators and a points scoring framework. The analysis applies the hierarchical model discussed in the preceding paper of this journal issue to case studies of waterfront areas and cultural quarters in three European cities: Belfast, Dublin and Barcelona. The evaluation permits performance comparisons to be made between the case studies regarding the sustainability of regeneration areas and projects, variations on an indicator set basis and the sensitivity of scores. Conclusions are drawn concerning regeneration practice, the extent to which sustainability principles are adhered to, potential policy benefits and the applicability of the model.”

Hermans, C. et al. 2007, “Collaborative environmental planning in river management: An application of multicriteria decision analysis in the White River Watershed in Vermont” in Journal of Environmental Management, vol. 84, issue 4, pp. 534-546
HAVE FULL TEXT
Multicriteria decision analysis (MCDA) provides a well-established family of decision tools to aid stakeholder groups in arriving at collective decisions. MCDA can also function as a framework for the social learning process, serving as an educational aid in decision problems characterized by a high level of public participation. In this paper, the framework and results of a structured decision process using the outranking MCDA methodology preference ranking organization method of enrichment evaluation (PROMETHEE) are presented. PROMETHEE is used to frame multi-stakeholder discussions of river management alternatives for the Upper White River of Central Vermont, in the northeastern United States. Stakeholders met over 10 months to create a shared vision of an ideal river and its services to communities, develop a list of criteria by which to evaluate river management alternatives, and elicit preferences to rank and compare individual and group preferences. The MCDA procedure helped to frame a group process that made stakeholder preferences explicit and substantive discussions about long-term river management possible

Hezri, A. & Dovers, S. 2009, “Australia’s Indicator-Based Sustainability Assessments and Public Policy” in Australian Journal of Public Administration, vol. 68, issue 3, pp. 303-318
HAVE ABSTRACT
Operationalising sustainability has proven difficult because the concept is ambiguous and fraught with contradictions. In response, sustainability indicator systems to characterise and measure sustainable development have been developed globally. The focus of this article is on how indicators function as an instrument of policy to enhance achievement of environmental and sustainability policy in Australia at national scale. In theory, by integrating information from the environmental, social and economic domains and then feeding knowledge to a wide range of policy sectors, sustainability indicator systems may facilitate cross-agency and -portfolio connectivity. Key characteristics of the various sustainability indicator systems developed to inform policy at the national level are described, noting the diverse approaches to indicator development in Australia. The structural relevance of indicator systems to the mechanisms of policy and institutions in Australia is discussed, and corresponding issues of institutional fit are analysed

Ioris, A. et al. 2008, “The development and application of water management sustainability indicators in Brazil and Scotland”, in Journal of Environmental Management, vol. 88, issue 4, pp. 1190-1201
HAVE FULL TEXT
This paper reports the formulation and application of a framework of catchment-level water resource management indicators designed to integrate environmental, economic and social aspects of sustainability. The framework of nine indicators was applied to the River Dee and River Sinos catchments in Scotland and Brazil, respectively, following an indicator selection process that involved inputs from water management professionals in both countries, and a pilot exercise in Scotland. The framework was found to capture a number of key sustainability concerns, and was broadly welcomed by water resource managers and experts as a means of better understanding sustainable water resource management. Issues relating to poor water quality and public water supply were particularly prominent in the findings for the Sinos, while findings for the Dee suggested that more attention might be focused on building institutional capacity and public participation in catchment management. The use of some proxy indicators was required in both catchments due to poor data availability, and this problem may hinder the further development of indicator frameworks that attempt to better integrate environmental, economic and social dimensions of sustainability.

Janssen, M. & Ostrom. 2005, “Governing social-ecological systems” 1465-1509 in K. L. Juddand L. Tesfatsion, (eds) Handbook of computational economics II: agent-based computational economics. Elsevier, Amsterdam, The Netherlands
BOOK, LIBRARY DOES NOT HAVE

Janssen, M. et al. 2006, “Toward a Network Perspective of the Study of Resilience in Social-Ecological Systems”, in Ecology and Society, vol. 11, no. 1, p. 15-
HAVE FULL TEXT
Formal models used to study the resilience of social-ecological systems have not explicitly included important structural characteristics of this type of system. In this paper, we propose a network

perspective for social-ecological systems that enables us to better focus on the structure of interactions between identifiable components of the system. This network perspective might be useful for developing formal models and comparing case studies of social-ecological systems. Based on an analysis of the case studies in this special issue, we identify three types of social-ecological networks: (1) ecosystems that are connected by people through flows of information or materials, (2) ecosystem networks that are disconnected and fragmented by the actions of people, and (3) artificial ecological networks created by people, such as irrigation systems. Each of these three archytypal social-ecological networks faces different problems that influence its resilience as it responds to the addition or removal of connections that affect its coordination or the diffusion of system attributes such as information or disease

Kemp, R., Parto, S. & Gibson, R. 2005, "Governance for sustainable development: moving from theory to practice", in International Journal of Sustainable Development, vol. 8, no. 1/2, pp. 12-30
HAVE FULL TEXT
In this paper we examine and elaborate on the central elements of sustainable development and governance, considering their interrelations as they have emerged from the core themes in sustainable development discourses over the past decade and a half. We argue that sustainability is best viewed as a socially instituted process of adaptive change in which innovation is a necessary element. We discuss four key elements of governance for sustainability, which are integrated into the concept of transition management. The result is a conceptual framework for policy-making and action-taking aimed at progress towards sustainability

Kilvington, M. 2007, "Social learning as a framework for building capacity to work on complex environmental management problems"
http://www.landcareresearch.co.nz/publications/researchpubs/Social_learning_review.pdf
HAVE FULL TEXT
Social learning is emerging as a useful framework to support collective decision making and action. This on-line article provides a concise review of the growing body of literature on social learning. It highlights how social learning is supported by a number of different elements. These elements can be broken down into three clusters: (1) learning and thinking; (2) group participation and interaction; and (3) social and institutional

Koo, D. & Ariaratnam, S. 2008, "Application of a sustainability model for assessing water main replacement options" in Journal of Construction Engineering and Management, vol. 134, issue 8, pp. 563-574
HAVE FULL TEXT
Sustainable development, conceived as a new and multidisciplinary paradigm, is receiving much attention throughout the global community. The purpose of this paper is to apply the sustainability assessment model (SAM), an assessment and decision making methodology, to a water main replacement project in an urban environment to determine the most sustainable project alternative among three possible options. This case study presents the use of SAM in considering various multicriteria sustainability indicators while working towards achieving sustainability enhancement. Objectives of sustainability enhancement include: (1) minimizing environmental impact; (2) maximizing economical benefit and output; (3) social and cultural conservation and promotion; and (4) satisfying basic requirements such as structural soundness and capacity. Six assessment methods including the analytic hierarchy process, cost, pollution, energy, time estimation, and natural resource depletion analysis are used for both qualitative and quantitative sustainability indicators. The weighted sum model is then utilized to integrate the six independent assessment results to elicit the final decision

Lai, E., Lundie, S. & Ashbolt, N. 2008, "Review of multi-criteria decision aid for integrated sustainability assessment of urban water systems" in Urban Water Journal, vol. 5, issue 4, pp. 315-327
HAVE FULL TEXT
Integrated sustainability assessment is part of a new paradigm for urban water decision making. Multi-criteria decision aid (MCDA) is an integrative framework used in urban water sustainability assessment, which has a particular focus on utilising stakeholder participation. Here MCDA is reviewed in the context of urban water management used in a decision making framework. Three other commonly used integrated approaches in urban water management (cost-benefit analysis, triple bottom line and integrated assessment) are compared with MCDA. Generic types of shortcomings associated with MCDA are discussed to provide an understanding of MCDA's limitation in urban water

management decision making; including 1) preferential independency, 2) double counting and under-counting, and 3) transparency of MCDA methods and results

Lebel, L. et al 2006, “Governance and the capacity to manage resilience in regional social-ecological systems” in Ecology and Society 11(1):19. [online] URL:
http://www.ecologyandsociety.org/vol11/iss1/art19/
HAVE FULL TEXT
The sustainability of regional development can be usefully explored through several different lenses. In situations in which uncertainties and change are key features of the ecological landscape and social organization, critical factors for sustainability are resilience, the capacity to cope and adapt, and the conservation of sources of innovation and renewal. However, interventions in social-ecological systems with the aim of altering resilience immediately confront issues of governance. Who decides what should be made resilient to what? For whom is resilience to be managed, and for what purpose? In this paper we draw on the insights from a diverse set of case studies from around the world in which members of the Resilience Alliance have observed or engaged with sustainability problems at regional scales. Our central question is: How do certain attributes of governance function in society to enhance the capacity to manage resilience? Three specific propositions were explored: (1) participation builds trust, and deliberation leads to the shared understanding needed to mobilize and self-organize; (2) polycentric and multilayered institutions improve the fit between knowledge, action, and social-ecological contexts in ways that allow societies to respond more adaptively at appropriate levels; and (3) accountable authorities that also pursue just distributions of benefits and involuntary risks enhance the adaptive capacity of vulnerable groups and society as a whole. Some support was found for parts of all three propositions. In exploring the sustainability of regional social-ecological systems, we are usually faced with a set of ecosystem goods and services that interact with a collection of users with different technologies, interests, and levels of power. In this situation in our roles as analysts, facilitators, change agents, or stakeholders, we not only need to ask: The resilience of what, to what? We must also ask: For whom?

Listowski, A. et al. 2009, in “Concepts towards a novel integrated assessment methodology of urban water reuse”, in Desalination and Water Treatment, vol 11. pp.81-92
HAVE FULL TEXT
Traditional supplies of large volumes of water and wastewater disposal technologies have offered a linear solution, thus intensifying environmental stress. In addition, provision of urban infrastructure especially any major augmentations are often the impractical or economically prohibitive. Urban water cycle should be viewed as an interactive and coordinated approach involving:
- Available water resources,
- Appropriate treatment technology producing fit for purpose water quality, and
- Ascertaining long term balance between environmental, social and economic issues.
Implementation of integrated water reuse scheme requires major paradigm shift within a number of technical (science, technology and knowledge) and non-technical (socio-cultural, economic, environment) dimensions. There are number of questions that arise from this scenario:
- How to make decisions regarding selection, design, implementation and operation of any recycling scheme?
- What knowledge is necessary to improve decision-making process?
- How to assess and compare performance of the scheme?
- What are the parameters for uniform evaluation process?
The aim of this paper is to introduce a new concept for integrated assessment methodology that can be applied for urban water reuse schemes. The conceptual assessment methodology relies on decision that treatment technology represents a leading theme and is supported by selection of socio-economic and environmental factors, thus enabling holistic evaluation and quantification of the outcomes.

Lundehn, C. & Morrison, G. 2007, “An assessment framework for urban water systems – a new approach combining environmental systems with service supply and consumer perspectives” in Highway and Urban Environment, vol. 12, pp. 559-577
HAVE ABSTRACT
There is increasing awareness that improvements to urban water systems (UWS) should be based on an assessment of sustainability that safeguards the well-being of both humans and ecosystems. There is a need for coherent and participatory frameworks, which integrate existing assessment tools. A

combined framework for UWS assessment is presented here and tested by application in a case study in Accra (Ghana). The approach combines the results of interviews with, and questionnaires to consumers and stakeholders with data collection of environmental sustainability indicators (ESIs) for the UWS. Safe water quality, safe access to water, and affordable water were identified as key criteria in the Accra dialogues, while healthy ecosystems and reducing environmental effects were considered least important. The ESI results reveal the need for an increased awareness and monitoring if the requirements of and challenges for a satisfactory UWS are to be met. The approach expands from existing methodologies towards a greater involvement of the general public and thereby provides a link between environmental sustainability and the service supply, and consumer perspectives

Maheephala, S. et al. 2006, "Assessing water service provision scenarios using the concept of sustainability", 2nd IWA Leading Edge Conference on Sustainability, IWA Publishing, London.
BOOK – SOME PAGES AVAILABLE ONLINE, COUNDN'T COPY ABSTRACT, EXCEPT FOM ABSTRACT BELOW James wants; can't find
"…This paper addresses this challenge and presents a method for the assessment of alternative urban water servicing options based on the concept of sustainability. The proposed assessment method uses multi-criteria assessment method to integrate social, economic and environmental aspects of decision-making. It is applicable at the planning and conceptual design stages of the water system of Greenfield and infill development schemes. Briefly, it requires a set of sustainability assessment criteria and at least one measurable indicator for each criterion be defined, values of indicators to be estimated for alternative scenarios and ranking of scenarios using a multi-criteria assessment method. Applicability of the proposed method has been demonstrated using 750 hectares of Greenfield development in Papamoa in New Zealand."
http://books.google.co.nz/books?hl=en&lr=&id=ajA4ZN-KIvYC&oi=fnd&pg=PA25&dq=method+integrate+OR+integration+OR+integrative+%22sustainability+assessment%22+-climate+-medical&ots=xseo3iIdrM&sig=UE7UKIKF9xq1yGc1gwIAbwnM5Sk#v=onepage&q=method%20integrate%20OR%20integration%20OR%20integrative%20%22sustainability%20assessment%22%20-climate%20-medical&f=false

Mendoza, G. & Prabhu, R. 2003, "Qualitative multi-criteria approaches to assessing indicators of sustainable forest resource management" in Forest Ecology and Management, vol. 174, pp. 329–343.
HAVE FULL TEXT
Criteria and indicators (C&I) have become primary tools in implementing the principle of sustainable resource management. To carry out this principle, it is necessary to develop methodologies that can holistically and systematically generate relevant indicators for a particular forest or resource management unit. This paper describes some methodologies that can be used as tools to carry out structured analysis of C&I. Multi-criteria analysis (MCA) is used as a decision-making tool to analyze and evaluate multiple C&I under a participatory group decision-making environment. Use of the method enables the generation of C&I, estimation of their relative importance, estimation of the performance of each indicator relative to its desired condition, and assessment of the indicators' combined effect or impact. In addition to generating C&I and estimating their relative importance and performance, the paper also presents a soft methodology, called cognitive mapping, which can be used to assess the crossindicator interaction, linkages, and connectivities of the indicators. The method attempts to evaluate the overall cumulative impacts of all indicators, individually and collectively, as they impact sustainability directly and indirectly through their interactions with other indicators

Mendoza, G. & Martins, 2006, "Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms" in Forest Ecology and Management, vol. 230, issue: 1-3, pp. 1-22
HAVE FULL TEXT
Multi-criteria decision analysis (MCDA) is an umbrella approach that has been applied to a wide range of natural resource management situations. This paper has two purposes. First, it aims to provide a critical review of MCDA methods applied to forest and other natural resource management. The review seeks to layout the nature of the models, their inherent strengths and limitations. Models are categorized based on different classification schemes and are reviewed by describing their general characteristics, approaches, and fundamental properties. The review goes beyond traditional MCDA

techniques; it describes new modelling approaches to forest management. The second purpose is to describe new MCDA paradigms aimed at addressing the inherent complexity of managing forest ecosystems, particularly with respect to multiple criteria, multi-stakeholders, and lack of information. Comments about, and critical analysis of, the limitations of traditional models are made to point out the need for, and propose a call to, a new way of thinking about MCDA as they are applied to forest and natural resource management planning. These new perspectives do not undermine the value of traditional methods; rather they point to a shift in emphasis—from methods for problem solving to methods for problem structuring

Milman, A. & Short, A. 2008, “Incorporating resilience into sustainability indicators: An example for the urban water sector” in Global Environmental Change-Human and Policy Dimensions, vol. 18, issue 4, pp. 758-767
HAVE FULL TEXT
The development and use of indicators is common practice in efforts to promote urban sustainability. Indicators used to measure urban sustainability tend to focus narrowly on describing the current state of the urban system. Although a time series analysis using these indicators may lend insights into trends towards or away from certain 'sustainability' goals, existing indicators of urban sustainability do not provide information on the ability or the likelihood that the current system state can be maintained or improved over time. Indicators that incorporate a measure of system resilience would provide useful information on system sustainability. Through development of a new indicator, Water Provision Resilience (WPR), we provide an example of how measures of resilience could be incorporated into sustainability indicators. The new indicator adds six color codings to the existing indicator 'percent of the population with access to safe water.' Each color coding represents a measure of the ability of the water system to maintain or improve the current percent of the population with access to safe water in key areas of the water provision sector: supply, infrastructure, service provision, finances, water quality and governance. The metric is then applied to three cities. The goal in developing this metric is to provide a starting point for re-thinking the metrics used to measure progress and sustainability in order to incorporate the ability to absorb and adapt to stresses into sustainability analysis

Moctezuma-Malagon, A. et al. 2008, “A methodology for evaluating the sustainability of inland wetland systems” in Aquaculture International, vol. 16, issue 6, pp. 525-537
HAVE FULL TEXT
Sustainable development is a process in which present and future human needs can be satisfied without degrading the socio-environmental systems that they rely upon. Inland wetlands are diverse ecosystems that provide numerous goods and services. Current wetland evaluation methodologies include inventories, monitoring of hydrological and climatic variables, population dynamic studies, mechanistic models, the use of geographical information systems, and measuring of biotic indices. Sustainability evaluation methods, which combine in a systemic approach production efficiency, ecosystem function, environmental services, and social values are required. In this study, a methodological framework was developed and applied to a case in the High Lerma River Basin in Central Mexico. Criteria and indicators comprising the environmental, economic, and social dimensions, as well as systemic attributes such as productivity, stability, resilience, self-reliance, and equity were derived and measured to evaluate wetland management systems. Native species were compared to the predominant carp monoculture system. Results from the evaluated indicators showed that the native species multicultural management system can be a feasible, profitable, and self-reliant conservation and utilization alternative. Some difficulties were faced in finding reliable estimates to be used as reference values, and potential indicators for future, long-term evaluations are proposed. The application of sustainability indicators to wetland management proved to be an objective and trustworthy methodological alternative, which could increase both wetland sustainability and the continuous development of integrated assessment tools for its evaluation

O’Conner, M. 2007, Deliberative sustainability assessment with the on-line kerDST deliberation support tool, http://www.ferloo.com/IMG/pdf/DSDV-FIC-IACAD-C3ED-20070403-00024.pdf
HAVE FULL TEXT
This article presents the design principles and application of the on-line deliberation support tool KERDST (developed by the KerBabel™ team at the C3ED) for sustainability assessment, in thencontext of the European SRDTOOLS project concerned with assessing regional development projects and policies relative to sustainability criteria. After a succinct exposition of the ‘problem of social choice’, we introduce the ‘Deliberation Matrix’ as a framework for multi-criteria multi-stakeholder

assessment and then present the KERDST system for managing and mobilising indicators. To conclude, we highlight the design principle of a “representative diversity of indicators” relative to the (impossible) ideal of monetary cost-benefit analysis of an inventory of project costs and benefits

Olalla-Tarraga, MA. 2006, “A conceptual framework to assess sustainability in urban ecological systems” in International Journal of Sustainable Development & World Ecology, vol. 13, issue 1, pp. 1-15
HAVE FULL TEXT
Contemporary approaches in urban ecology should take into account interactions and feedbacks between socio-ecological dimensions. Sustainability indicators are key tools to assess such integration, but initiatives are diverse and there is no agreed framework for the assessment of such interactions and feedbacks. Despite this formidable challenge, several attempts have been made to reach out beyond the traditional subject areas of environment, economy and society in frameworks for indicators of sustainable development. While efforts to develop such a methodology have been made from different scales and levels of analyses, as we descend to the local scale, initiatives are multiplied. There is a need to channel the diversity of these various initiatives and standardize some concepts and methods. In this paper I deal with the pros and cons of the most commonly applied conceptual frameworks, looking for a method able to optimize the main purposes of sustainable development indicators. Starting from this foundation, I build a new proposal combining different methods previously applied in some research fields. A single Hierarchical Framework, developed to assess sustainable forest management is represented graphically through a visual model presented in the Johannesburg Summit of 2002, the Dashboard of Sustainability. An appraisal of the local contribution to global sustainability is obtained by introducing the Ecological Footprint into the methodological procedure.

Olsson, J. et al. 2009, “A goal oriented indicator framework to support integrated assessment of new policies for agri-environmental systems” in Environmental Science and Policy, vol. 12, issue 5, pp. 562-572
HAVE FULL TEXT
The goal oriented framework (GOF) for indicators has been developed as part of a comprehensive research project developing computerised tools for integrated assessment of the effects of new policies or technologies on agricultural systems (SEAMLESS-IF). The ambition has therefore been to create an indicator framework where the environmental, economic and social dimensions of sustainable development can be related to each other in a consistent way. Integrated assessment tools rely on such frameworks to capture and visualise trade-offs (antagonisms or synergies) among indicators between and within the three dimensions of sustainable development. The specific aims of this paper are to (i) present the GOF (ii) present how the GOF can be used to select indicators within the integrated assessment framework SEAMLESS-IF and (iii) discuss the advantages and limitations with the proposed approach. We show that the GOF has several advantages. Its major rewards are its relative simplicity and the possibility to link indicators to policy goals of each dimension of sustainability and thereby facilitate the comparison of the impacts of the new policy on the different dimensions. Another important feature of the GOF is its multi-scale perspective, which will enable the comparison of effects of a new policy between scales. Yet, as typical for all indicator frameworks, the GOF has also biases either instigated by the issues the included models cover or by the stakeholders¿ selection of indicators. However, due to the way the GOF and its indicators are technically implemented in SEAMLESS-IF, it can easily be extended and include new indicators to increase and update its policy relevance

Organisation for Economic Co-operation and Development. 2008, “Handbook on Constructing Composite Indicators: Methodology and User Guide”, Organisation for Economic Co-operation and Development (OECD): France.
HAVE FULL TEXT

O’Riordan, J. 2006, “Governance for sustainability in water resource management: the role of collaborative decision-making”, 2nd IWA Leading Edge Conference on Sustainability, IWA Publishing, London.
BOOK – SOME PAGES AVAILABLE ONLINE, COUNDN’T COPY ABSTRACT, LOOKS INTERESTING… James wants; can’t find

http://books.google.co.nz/books?hl=en&lr=&id=ajA4ZN-KIvYC&oi=fnd&pg=PA25&dq=method+integrate+OR+integration+OR+integrative+%22sustainability+assessment%22+-climate+-medical&ots=xseo3iIdrM&sig=UE7UKIKF9xq1yGc1gwIAbwnM5Sk#v=onepage&q=method%20integrate%20OR%20integration%20OR%20integrative%20%22sustainability%20assessment%22%20-climate%20-medical&f=false

Osterblom, H. et al. 2010, “Making the ecosystem approach operational – Can regime shifts in ecological and governance systems facilitate the transition?” in Marine Policy, In Press
HAVE FULL TEXT
Effectively reducing cumulative impacts on marine ecosystems requires co-evolution between science, policy and practice. Here, long-term social–ecological changes in the Baltic Sea are described, illustrating how the process of making the ecosystem approach operational in a large marine ecosystem can be stimulated. The existing multi-level governance institutions are specifically set up for dealing with individual sectors, but do not adequately support an operational application of the ecosystem approach. The review of ecosystem services in relation to regime shifts and resilience of the Baltic Sea sub-basins, and their driving forces, points to a number of challenges. There is however a movement towards a new governance regime. Bottom-up pilot initiatives can lead to a diffusion of innovation within the existing governance framework. Top-down, enabling EU legislation, can help stimulating innovations and re-organizing governance structures at drainage basin level to the Baltic Sea catchment as a whole. Experimentation and innovation at local to the regional levels is critical for a transition to ecosystem-based management. Establishing science-based learning platforms at sub-basin scales could facilitate this process

Pahl-Wostl, C. 2002. Towards sustainability in the water sector: the importance of human actors and processes of social learning. Aquatic Sciences 64:394-411.
HAVE FULL TEXT
Current regimes in resource management are often unsustainable as judged by ecological, economic and social criteria. Many technological resource management regimes are inflexible and not built to adapt to changes in environmental, economic or social circumstances. This inflexibility poses problems in a world characterized by fast change. The water sector is currently undergoing major processes of transformation at local, regional and global scales. Today’s situation is challenged by uncertainties, e.g., in water demand (diminishing in industrialized countries, rising in developing countries), by worsening water quality, by pressure for cost-efficient solutions, and by fast changing socio-economic boundary conditions. One expects additional uncertainties, due to climate change, such as a shift in the pattern of extreme events. Hence, new strategies and institutional arrangements are required to cope with risk and change in general. When one considers processes of transformation and change, the human dimension is of particular importance. Institutions and rule systems may cause resistance to change but can also enable and facilitate necessary transformation processes. This paper explores conceptual approaches in social learning and adaptive management. It introduces agent-based modelling, and the link between analytical modelling and participatory approaches as promising new developments to explore and foster changes towards sustainability and the required transformations in technological regimes and institutional settings

Pahl-Wostl, C. 2006, “The Importance of Social Learning in Restoring the Multifunctionality of Rivers and Floodplains, in Ecology and Society, vol. 11, no. 1, pp. 10-
HAVE FULL TEXT
Those involved in floodplain restoration have to cope with historical conflicts between human and ecosystem needs. The topic is of high importance in Europe due to the European Water Framework Directive that requires restoration and/or maintenance of a “good ecological status of aquatic ecosystems.” However, the seeming trade-off between flood protection and floodplain restoration may change due to a shift in the water management paradigm toward more integrated approaches, in contrast to the command and control approach of the past. This shift in paradigm is summarized in the guiding principle for water management in the Netherlands “Living with floods and give room to water” rather than “Fighting against water.” The paper discusses the role of social learning in the transition toward the adaptive management of floodplains and rivers that is required to restore and maintain multifunctional riverine landscapes. In addition to the uncertainties resulting from our limited knowledge about the complex spatiotemporal dynamics of floodplains, we have to take into account the ambiguities that arise as a result of the different perceptions of stakeholders.

Pahl-Wostl, C., M. Craps, E. Mostert, D. Tabara, and T. Taillieu. 2007, "Social learning and water resources management" in Ecology and Society, vol. 12, no.1, pp. 5-
HAVE FULL TEXT
Natural resources management in general, and water resources management in particular, are currently undergoing a major paradigm shift. Management practices have largely been developed and implemented by experts using technical means based on designing systems that can be predicted and controlled. In recent years, stakeholder involvement has gained increasing importance. Collaborative governance is considered to be more appropriate for integrated and adaptive management regimes needed to cope with the complexity of social-ecological systems. The paper presents a concept for social learning and collaborative governance developed in the European project HarmoniCOP (Harmonizing Collaborative Planning). The concept is rooted in the more interpretive strands of the social sciences emphasizing the context dependence of knowledge. The role of frames and boundary management in processes of learning at different levels and time scales is investigated. The foundation of social learning as investigated in the HarmoniCOP project is multiparty collaboration processes that are perceived to be the nuclei of learning processes. Such processes take place in networks or "communities of practice" and are influenced by the governance structure in which they are embedded. Requirements for social learning include institutional settings that guarantee some degree of stability and certainty without being rigid and inflexible. Our analyses, which are based on conceptual considerations and empirical insights, suggest that the development of such institutional settings involves continued processes of social learning. In these processes, stakeholders at different scales are connected in flexible networks that allow them to develop the capacity and trust they need to collaborate in a wide range of formal and informal relationships ranging from formal legal structures and contracts to informal, voluntary agreements

Pahl-Wostl, C. 2009, "A conceptual framework for analysing adaptive capacity and multi-level learning processes in resource governance regimes" in Global Environmental Change, vol. 19, issue 3, pp. 354-365
HAVE FULL TEXT
Governance failures are at the origin of many resource management problems. In particular climate change and the concomitant increase of extreme weather events has exposed the inability of current governance regimes to deal with present and future challenges. Still our knowledge about resource governance regimes and how they change is quite limited. This paper develops a conceptual framework addressing the dynamics and adaptive capacity of resource governance regimes as multi-level learning processes. The influence of formal and informal institutions, the role of state and non-state actors, the nature of multi-level interactions and the relative importance of bureaucratic hierarchies, markets and networks are identified as major structural characteristics of governance regimes. Change is conceptualized as social and societal learning that proceeds in a stepwise fashion moving from single to double to triple loop learning. Informal networks are considered to play a crucial role in such learning processes. The framework supports flexible and context sensitive analysis without being case study specific. First empirical evidence from water governance supports the assumptions made on the dynamics of governance regimes and the usefulness of the chosen approach. More complex and diverse governance regimes have a higher adaptive capacity. However, it is still an open question how to overcome the state of single-loop learning that seem to characterize many attempts to adapt to climate change. Only further development and application of shared conceptual frameworks taking into account the real complexity of governance regimes can generate the knowledge base needed to advance current understanding to a state that allows giving meaningful policy advice

Perrings, C. 2006, "Resilience and sustainable development" in Environment and Development Economics, vol. 11, issue 4, pp. 417-427
NO ABSTRACT, HAVE FULL TEXT

Pinter, L., Hardi, P. & Bartelmus, P. 2005, Sustainable Development Indicators. Proposals for a Way Forward, International Institute for Sustainable Development
HAVE FULL TEXT
This paper was prepared for the United Nations Division for Sustainable Development (UNDSD) expert meeting (New York, 13-15 December 2005) on sustainable development indicators (SDIs). The paper provides a review of key achievements and emerging trends in the field of SDIs, reflects on the

role of the indicator system developed by the United Nations Commission for Sustainable Development (UNCSD) and offers a set of options and suggestions for the way forward

Plummer, R. & Armitage, D. 2006, “A resilience-based framework for evaluating adaptive co-management: Linking ecology, economics and society in a complex world” in Ecological Economics, vol. 61, issue 1, pp. 62-74
HAVE FULL TEXT
Adaptive co-management brings together collaborative and adaptive approaches in pursuit of sustainable resource use and social–ecological resilience. Enthusiasm for this management approach, however, is countered by recent critiques regarding outcomes. A lack of evidence from consistent evaluation of adaptive co-management further exacerbates this situation. This paper revisits the issue of evaluation in natural resource management and recasts it in light of complex adaptive systems thinking. An evaluative framework for adaptive co-management is developed which directs attention toward three broad components: ecosystem conditions, livelihood outcomes and process and institutional conditions. Scale-specific parameters are offered for each component to facilitate systematic learning from experience and encourage cross-site comparisons. Conclusions highlight the importance of systematically incorporating evaluation into the adaptive co-management process and recognize the challenge for resource agencies and researchers to shift from a conventional to a complex adaptive system perspective.

Pope, J. & Grace, W. 2006a, “Sustainability Assessment in context: Issues of process, policy and governance” in Journal of Environmental Assessment Policy and Management, vol. 8, no. 3, pp. 373-398.
CAN’T FIND

Proctor, W. & Qureshi M. 2005. Multi-criteria Evaluation Revisited, Proceedings of the Australia New Zealand Society for Ecological Economics Conference.
CAN’T FIND

Proctor W, Drechsler M 2006. Deliberative multicriteria evaluation. Environment and Planning C: Government and Policy 24: 169-190.
HAVE FULL TEXT
Multicriteria evaluation (MCE) is a well-tried and effective procedure for structuring and aiding complex decisionmaking processes, especially those involving environmental considerations. Formal deliberative processes have also been successful in aiding understanding and meeting consensus in complex and difficult decision problems which involve more than one decisionmaker. Here, both approaches are combined in a new technique called ‘deliberative multicriteria evaluation’ to assist a group of natural resource managers to decide on a suitable option for recreation and tourism activities in the upper Goulburn – Broken Catchment of Victoria, Australia. This approach is an attempt to combine the advantages of MCE, providing structure and integration in complex decision problems, with the advantages of deliberation and stakeholder interaction provided by a ‘citizens’ jury’. An important outcome of the process was the discovery of some crucial aspects of the decision problem that required deeper understanding and assessment if that preferred strategy were to have the desired results. Some suggestions for improving the process are provided but, in general, the stakeholder jury was regarded as a helpful and useful procedure by the decisionmakers and one which aided them in their understanding of the issues of a complex decisionmaking problem

Proctor, W. 2009, “Multi-criteria Analysis” in Argyrous G ed. Evidence for policy and decision-making: A practical guide, Sydney, Australia, UNSW Press.
CAN’T FIND, NOT IN LIBRARY

Reed, M. et al, 2006, “An adaptive learning process for developing and applying sustainability indicators with local communities” in Ecological Economics, vol. 59, pp. 40-418
HAVE FULL TEXT
Sustainability indicators based on local data provide a practical method to monitor progress towards sustainable development. However, since there are many conflicting frameworks proposed to develop indicators, it is unclear how best to collect these data. The purpose of this paper is to analyse the literature on developing and applying sustainability indicators at local scales to develop a methodological framework that summarises best practice. First, two ideological paradigms are

outlined: one that is expert-led and top–down, and one that is community-based and bottom–up. Second, the paper assesses the methodological steps proposed in each paradigm to identify, select and measure indicators. Finally, the paper concludes by proposing a learning process that integrates best practice for stakeholder-led local sustainability assessments. By integrating approaches from different paradigms, the proposed process offers a holistic approach for measuring progress towards sustainable development. It emphasizes the importance of participatory approaches setting the context for sustainability assessment at local scales, but stresses the role of expert-led methods in indicator evaluation and dissemination. Research findings from around the world are used to show how the proposed process can be used to develop quantitative and qualitative indicators that are both scientifically rigorous and objective while remaining easy to collect and interpret for communities

Roux, D., Murray, K. & van Wyk, E. 2007, “Principles enabling learning environments for good ecosystem governance” in Governance as a Trialogue: Government-Society-Science in Transition, pp.253-280
HAVE ABSTRACT
Complexity and, by implication, change and uncertainty, are inherent features of ecosystems. In managing ecosystems, or linked social-ecological systems, decisions are often based on insufficient or uncertain data and information. Appropriate and sufficient knowledge, which essentially resides in people, is a critical factor for making informed decisions under such circumstances. Informed action is a function of what we know, and our knowledge is a product of what and how we have learned. Because of the central importance of learning, this chapter proposes that the development of an appropriate learning capability should not be left to chance but should be the result of deliberate intervention to establish the conditions for an organisation to operate in a learning mode. Focusing on organisations or agencies with mandates for ecosystem governance, the chapter sets out to identify the principles that will enable the creation of such learning environments. Firstly, the key concepts of knowledge, learning and ecosystem governance are defined. Secondly, the chapter identifies main issues of concern regarding (a) the type of knowledge that needs to be created for good ecosystem governance; (b) the desirable processes for learning or knowledge creation; and (c) the characteristics of good learners. Thirdly, these main issues (10 in all) form the basis for formulating nine principles intended to enable the setting up of appropriate learning environments for ecosystem governance. The proposed principles are summarised as follows. Good ecosystem governance requires positively persistent and adaptive people with a culture of empathy for other knowledge systems and levels. Their knowledge must be trans-disciplinary, moulded by a common future focus, acquired by patiently engaging their prior knowledge and learning by doing, in an environment of social knowledge sharing. It is concluded that good learning practice would promote the achievement of some of the principles underlying the practice of good ecosystem governance, notably effective stakeholder engagement, adaptability and transformability. The proposed learning principles could be used as a framework to assess the learning proficiency of ecosystem management agencies and to develop learning strategies for such agencies
Salgado, P. et al. 2009, “Participative multi-criteria analysis for the evaluation of water governance alternatives. A case in the Costa del Sol (Malaga)” in Ecological Economics, vol. 68, issue 4, pp. 990-1005
HAVE FULL TEXT
This paper presents the results of a participatory multi-criteria evaluation of urban water supply alternatives in Costa del Sol Occidental in Málaga (Spain). Multi-criteria and social research techniques were combined, relying on social actors' involvement throughout the research work. The paper documents that participatory processes allowed to unveil framings, perspectives, values and interests, as well as understanding of the social and institutional context in which the waters are governed. The deployment of this methodological approach has in this case provided an extended diagnosis which was the starting point to explore non a priori foreseen management alternatives by the researchers. The methodological approach proved useful for problem structuring in a collective, flexible and iterative way, unveiling existing water management conflicts and their motivations, and improving the quality and effectiveness of the information exchange and the reflection process, among social actors, even if the whole process developed in a research setting rather than in a policy for real one. Participatory multi-criteria evaluation could in this case, contribute to identify policy options that could be better defended before all the social actors, including the general public, thus attempting to achieve shared ground. The paper suggests some methodological considerations for applying participatory multi-criteria evaluation based on the outcomes of the case study application

Scerri, A. & James, P. 2010, “Accounting for sustainability: combining qualitative and quantitative research in developing ‘indicators’ of sustainability” in International Journal of Social Research Methodology, vol. 13, issue 1, pp. 41-53
HAVE FULL TEXT
Indicators-based projects are currently central to many local, city-wide, national and international sustainability initiatives. The quantitative basis of many such projects means that achieving sustainability through them is often undertaken as a technical task. The size, scope and sheer number of indicators included within many such projects means that they are often unwieldy and resist effective implementation. Arguably, the techno-scientific 'edge' inherent in them tends to blur the possibilities for bringing into question the structures of power and criteria by which values are translated into practice. It limits the way that a community may use indicators to support sustainable practices or to challenge unsustainable practices. The article discusses some of the methodological issues that arise when setting out to develop and implement qualitative indicators of sustainability that incorporate some quantitative metrics. This alternative approach involves people in actively learning and negotiating over how best to put sustainability into practice. The aim of such a research method is to engage citizens in the job of achieving sustainability as a task of itself, undertaken on terms acceptable to them in the context of the communities in which they live

Scipiponi, A. et al, 2009, “The Dashboard of Sustainability to measure the local urban sustainable development: The case study of Padua Municipality” in Ecological Indicators, vol 9, issue 2, pp. 364-380
HAVE FULL TEXT
The adoption of suitable indicators is fundamental to implement sustainable development at the local level. It helps in the analysis and evaluation; supports the decisional process and helps the communication between the citizens and the society, in general. Furthermore, the aggregated indexes – by representing the observed context in a simple way – may help the community in the definition of effective improvement goals and also serve as important tools to monitor the fulfillment of the planned objectives. The Dashboard of Sustainability (DS) is a mathematical and graphical tool designed to integrate the complex influences of sustainability and support the decision-making process by creating concise evaluations. The city of Padua, Italy, agreed to use the DS in its Local Agenda 21 project, financed by the Italian Ministry of Environment. The available data were sufficient to design 61 useful indicators of environmental protection, economic development, and social promotion. The results of the analysis were discussed in Padua's community forum on the Local Agenda 21 process. The graphical and numerical results helped Padua reach a consensus on the plan for future sustainability, one that was understood and accepted by all the stakeholders. In order to adapt this tool to a local context, two changes in the methodologies were necessary: to measure urban sustainability, for which “ad hoc” set indicators were used; and to allow a comparative evaluation, for which the performances of Padua were evaluated over time. These changes were possible, thanks to the flexibility of such tool.

Singh, R. et al. 2009, “An overview of sustainability assessment methodologies”, in Ecological Indicators, vol. 9, pp. 189-212
HAVE FULL TEXT
“Sustainability indicators and composite index are increasingly recognised as a useful tool for policy making and public communication in conveying information on countries and corporate performance in fields such as environment, economy, society, or technological improvement. By visualizing phenomena and highlighting trends, sustainability indicators simplify, quantify, analyse and communicate otherwise complex and complicated information. There are number of initiatives working on indicators and frameworks for sustainable development (SD). This article provides an overview various sustainability indices applied in policy practice. The paper also compiles the information related to sustainability indices formulation strategy, scaling, normalisation, weighting and aggregation methodology.”

Spangenberg, J. 2009, “Sustainable development indicators: towards integrated systems as a tool for managing and monitoring a complex transition” in International Journal of Global Environmental Issues, vol. 9, no. 4, pp. 318-337
HAVE ABSTRACT, CAN’T FIND FULL TEXT
The situation for developing indicators is not clear-cut, but characterised by three dilemmata making it particularly challenging. Only by combining a strategic approach with short-term flexibility, the effort can be successful. These dilemmas are characterised as Indicators vs. Indicator Systems, Themes vs.

Systems and Governance insufficiencies. To address them, a systematic approach to sustainability indicators is suggested, based on first defining objectives and strategies in a systematic fashion. In a second step, indicators can be derived and boiled down to a set of headline indicators, as illustrated by a set of indicators suggestions resulting from several European research projects

Soma, K. 2010, "Framing participation with multicriterion evaluations to support the management of complex environmental issues" in Environmental Policy and Governance, vol. 20, issue 2, pp. 89-106
HAVE FULL TEXT
A repeated statement observed in the literature is that stakeholder participation can contribute to improve complex environmental management decisions. However, transparent and legitimate decision-making processes cannot be ensured without suitable involvement strategies and information treatments throughout the processes. The main goal in this study is to frame participatory processes with multicriterion evaluations to increase transparency of the decision support. The developed approach applies clearly defined roles of interest groups, experts and citizens, as well as alternatives presented on maps, criteria arranged in a hierarchy of decision elements and weights obtained by conducting deliberative processes with citizens. The approach is applied in a case study at municipal level in Norway to support coastal zone management decisions. Relevant interests and social values are systematically represented by the multicriterion evaluation framework utilized in the approach presented

Stringer, L. et al. 2006, "Unpacking "Participation" in the Adaptive Management of Social-Ecological Systems: a Critical Review", in Ecology and Society, vol. 11, no. 2, pp. -39
HAVE FULL TEXT
Adaptive management has the potential to make environmental management more democratic through the involvement of different stakeholders. In this article, we examine three case studies at different scales that followed adaptive management processes, critically reflecting upon the role of stakeholder participation in each case. Specifically, we examine at which stages different types of stakeholders can play key roles and the ways that each might be involved. We show that a range of participatory mechanisms can be employed at different stages of the adaptive cycle, and can work together to create conditions for social learning and favorable outcomes for diverse stakeholders. This analysis highlights the need for greater reflection on case study research in order to further refine participatory processes within adaptive management. This should not only address the shortcomings and successes of adaptive management as a form of democratic environmental governance, but should also unpack the links between science, institutions, knowledge, and power

Štreimikien?, D., Girdzijauskas, S. & Stoškus, L. 2009, "Sustainability Assessment Methods and Their Application to Harmonization of Policies and Sustainability Monitoring" in Environmental Research, Engineering and Management, vol. 48, no. 2. pp. 51-62
HAVE FULL TEXT
Sustainability assessment methods and techniques are analyzed in this paper. Criteria for sustainability assessment of state policies are defined as based on priorities of the National Sustainable Development Strategy. Technique for sustainability assessment of state policies is developed. An integrated indicators approach is constructed for monitoring implementation of a sustainable development strategy in Lithuania. This work aims at developing the technique for sustainability assessment of state policies and measures based on various sustainability assessment tools, methods and techniques developed by other scientists. The most important methods and techniques for sustainability assessment are analyzed and systematized. Based on the analysis of priorities of Lithuanian government policies the principal economic, environmental and social criteria for sustainability assessment of policies and measures are established for Lithuania. The proposed multi-criteria decision analysis technique is based on the integrated sustainability indices developed for the state, region, enterprise level and it includes a set of social, economic and environmental indicators of different levels. Monitoring technique for the progress achieved towards sustainability is applied to monitoring implementation of the Lithuanian sustainable development strategy

Tàbara, J. et al. 2008, "Integrated sustainability assessment of water systems: lessons from the Ebro River Basin" in International Journal of Innovation and Sustainable Development, Vol. 3, Nos. 1–2, pp.48–69.
HAVE FULL TEXT

This paper describes the co-development and implementation of visioning and experimenting exercises, agent-based modelling, and gaming tools in Integrated Sustainability Assessments (ISAs) involving stakeholders. These new tools are aimed at supporting reflexive learning and at building alternative policy-relevant knowledge and evaluative paradigms for managing sustainability. The specific case study relates to water management within the Ebro River Basin. Conclusions concern the use of these tools to represent complexity, to learn how conflict and collaboration between agents can be addressed, and to explore the roles played by power regimes, institutional rules, and culture in constraining or enhancing transition in the water domain

Taylor, A. & Fletcher, T. 2005, " 'Triple-Bottom-Line' assessment of urban stormwater projects" 10th International Conference on Urban Drainage, Copenhagen/Denmark, 21-26 August 2005
HAVE FULL TEXT
New guidelines have been developed and trialled in Australia to assist urban stormwater managers to assess options for projects that aim to improve urban waterway health. These guidelines help users to examine the financial, ecological and social dimensions of projects (i.e. the so-called 'triple-bottom-line'). Features of the assessment process described in the guidelines include use of multi criteria analysis, input from technical experts as well as nontechnical stakeholders, and provision of three alternative levels of assessment to suit stormwater managers with differing needs and resources. This paper firstly provides a background to the new guidelines and triple-bottom-line assessment. The assessment methodology promoted in the new guidelines is then briefly summarised. This methodology is compared and contrasted with European guidelines from the 'SWARD' project that have been primarily developed for assessing the relative sustainability of options involving urban water supply and sewerage assets. Finally, the paper discusses how assessment methodologies that evaluate the financial, ecological and social dimensions of projects can, under some circumstances, be used to evaluate the relative progress of options for urban water management on a journey towards the widely pursued, but vaguely defined goal of 'sustainable development'

Taylor, A. 2006, "Triple-bottom-line assessment of stormwater quality projects: advances in practicality, flexibility and rigour", in Urban Water Journal, vol. 3, issue 2, pp. 79-90
HAVE ABSTRACT James wants, can't find – but asked Cawthron
Decision-support frameworks that help users to consider financial, social and ecological costs and benefits (i.e., 'triple-bottom-line' decision frameworks) are increasingly being used by water managers. This paper describes the outcomes of a project that was undertaken to assist stormwater managers to complete triple-bottom-line assessments of proposed urban stormwater quality management measures. The primary outcomes have been: a set of guidelines that clearly define a twelve step assessment process; and the results of a trial that was used to test and refine the guidelines in Brisbane, Australia. A brief overview of the twelve step assessment process is provided. The strengths and weaknesses of typical triple-bottom-line assessment processes are highlighted through the trial, where seven alternative stormwater treatment and reuse options for new, medium density, residential estates were assessed and ranked to help inform broad policy and local design decisions. For managers of urban stormwater quality, we conclude that the guidelines are more tailored, practical and flexible than existing decision support frameworks, whilst maintaining a high degree of assessment rigour. The trial confirmed their value in building the capacity of local stakeholders and making more informed and carefully considered decisions.

United Nations. 2007, "Indicators of Sustainable Development: Guidelines and Methodologies", Third Edition, United Nations Publications, New York.
HAVE FULL TEXT
This publication presents the third set of Indicators of Sustainable Development and provides suggestions on how to adapt them to national conditions and priorities. It benefits from the active participation of and excellent collaboration with, a wide range of governments, international organizations, academic institutions, non-governmental organizations and individual experts. The indicators are a follow-up to the two earlier sets prepared under the work programme on indicators of sustainable development approved by the Commission on Sustainable Development in 1995. These earlier sets were published in 1996 and 2001. We hope that countries will find the publication useful whenever they are reviewing their existing indicators or developing new indicators to measure progress towards nationally defined goals for sustainable development. The indicators of sustainable development presented here reflect the valuable experiences of countries and international organizations over the past fifteen years since the adoption of Agenda 21 in Rio de Janeiro. With this

publication, we also hope to further the momentum at the national and international level to develop and apply sustainable development indicators. This will help the understanding of the various dimensions of sustainable development and their complex interactions and the facilitation of policy decisions aimed at achieving sustainable development goals.

van der Brugge, R. & van Raak, R. 2007, “Facing the Adaptive Management Challenge: insights from Transition Management” in Ecology and Society, vol. 12, no. 2, p. 33
HAVE FULL TEXT
Recent research suggests that transitions toward adaptive water management regimes are needed because current water management regimes cannot adequately respond to uncertainty. The pivotal question is how to understand and manage such transitions. The literature on adaptive management addresses this question in part, but must now move beyond the descriptive toward a prescriptive management framework. Transition management theory could help in meeting this challenge. The similarity of the theoretical starting points yet different applications offer fertile conditions for cross-pollination. We investigate three central concepts from the transition management literature for their potential contribution to adaptive management. In particular, the notions of arenas and shadow networks merit further study through joint research.

van Zeijl-Rozema, A. et al, 2008, “Governance for sustainable development: a framework” in Sustainable Development, vol. 16, issue 6, pp. 410-421
HAVE FULL TEXT
This paper explores how two complex concepts, governance and sustainable development, can be linked in order to obtain a better understanding of their interactions. We argue that the many perspectives on sustainable development can be mapped on the continuum between ecological sustainability and quality of life. Likewise, the variety of modes of governance can be captured between hierarchical governance and deliberative governance, depending on the degree of involvement of societal actors. From these two typologies we derive a framework for analysis of governance for sustainable development. We realize that the typology is a significant simplification of the complex debates about sustainable development and governance, but it might help scientists and policy makers to explore relevant dimensions of modes of governance for sustainable development and for setting a framework for empirical analysis. The main conclusion of this paper is that the debate on governance for sustainable development will be clarified if the perspective on sustainable development and the mode of governance for achieving it are made more explicit. Problems that are now exclusively associated with sustainable development might well be problems of governance for sustainable development. Without making explicit what type of governance for sustainable development is pursued, miscommunication between stakeholders and mismatches of the approach with the instruments used could be the result, thus hampering progress in implementing sustainable development

Walker, B. et al. 2004, “Resilience, adaptability and transformability in social-ecological systems” in Ecology and Society vol. 9, issue 2, pp. 5-
HAVE FULL TEXT

Wallis, A. et al, 2007, “Measuring regional sustainability: lessons to be learned”, in International Journal of Environment and Sustainable Development, vol. 6, no. 2, pp. 193-207
HAVE ABSTRACT James wants; can’t find
Developing tools for measuring progress toward sustainability has proven a challenging task. Indicators offer an excellent means to explore the success or otherwise of management strategies. They also allow reporting social, economic and environmental aspects of sustainability. To ensure that the tools developed are effective in measuring the progress toward sustainable futures, an evaluation of the methods and the indicators used must be undertaken so that with progress there is learning and with the new knowledge methods can be redesigned to better advance sustainability. This paper discusses a study carried out in the south west region of Victoria, Australia, using indicators as the basis for developing a tool to measure progress toward sustainability. By evaluating the methods and indicators used in the study this paper provides an insight into the challenges encountered and the lessons learned. Issues explored include selecting indicators, collating data, integrating social, economic and environmental aspects of sustainability and using an adaptive approach

Wiegand, J. 2010, “Assessment of temporal trends in ecosystem health using an holistic indicator” in Journal of Environmental Management, vol. 91, issue 7, pp. 1446-1445
HAVE FULL TEXT
The Ecosystem Approach to environmental management, with its explicit acknowledgment of the environmental, social and economic elements within ecological systems, is becoming increasingly adopted by managers and policy makers. However, there are few specific prescriptions as to exactly how these different elements should be integrated, or examples where historical changes have been tracked. Here, we assess the potential benefits and challenges of applying an Ecosystem Approach to management using one particular method, the Holistic Ecosystem Health Indicator (HEHI), which integrates data from the ecological, social and interactive dimensions into a single composite index of ecosystem 'health'. We apply it to one of the best-documented catchments in the UK, the Ythan Estuary in Aberdeenshire, Scotland, based on a 50-year time series of data. The evaluation of ecosystem health over time is lacking in previous applications of HEHI, yet is crucial for the assessment of the utility of this approach to ecosystem management…
Young, O. 2010, “Institutional dynamics: Resilience, vulnerability and adaptation in environmental and resource regimes” in Global Environmental Change, vol. 20, issue 3, pp. 378-385
HAVE FULL TEXT
Like all social institutions, governance systems that address human–environment relations – commonly know as environmental or resource regimes – are dynamic. Although analysts have examined institutional change from a variety of perspectives, a particularly puzzling feature of institutional dynamics arises from the fact that some regimes linger on relatively unchanged even after they have become ineffective, while others experience state changes or even collapse in the wake of seemingly modest trigger events. This article employs the framework developed to study resilience, vulnerability, and adaptation in socio-ecological systems (the SES framework) in an effort to illuminate the conditions leading to state changes in environmental and resource regimes. Following a discussion of several conceptual issues, it examines institutional stresses, stress management mechanisms, and the changes that occur when interactive and cumulative stresses overwhelm these mechanisms. An important conclusion concerns the desirability of thinking systematically about institutional reform in a timely manner, in order to be prepared for brief windows of opportunity to make planned changes in environmental regimes when stawte changes occur

Appendix 3. OECD Statistics Directorate Handbook on Constructing Composite Indicators: Methodology and User Guide.

Unclassified **STD/DOC(2005)3**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development **09-Aug-2005**

English - Or. English

STATISTICS DIRECTORATE

STD/DOC(2005)3
Unclassified

HANDBOOK ON CONSTRUCTING COMPOSITE INDICATORS: METHODOLOGY AND USER GUIDE

OECD Statistics Working Paper

by Michela Nardo, Michaela Saisana, Andrea Saltelli and Stefano Tarantola (EC/JRC)
Anders Hoffman and Enrico Giovannini (OECD)

English - Or. English

JT00188147

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

OECD STATISTICS WORKING PAPER SERIES

The OECD Statistics Working Paper Series - managed by the OECD Statistics Directorate - is designed to make available in a timely fashion and to a wider readership selected studies prepared by staff in the Secretariat or by outside consultants working on OECD projects. The papers included are of a technical, methodological or statistical policy nature and relate to statistical work relevant to the organisation. The Working Papers are generally available only in their original language - English or French - with a summary in the other.

Comments on the papers are welcome and should be communicated to the authors or to the OECD Statistics Directorate, 2 rue André Pascal, 75775 Paris Cedex 16, France.

The opinions expressed in these papers are the sole responsibility of the authors and do not necessarily reflect those of the OECD or of the governments of its Member countries.

http://www.oecd.org/std/research

ABSTRACT

This Handbook aims to provide a guide for constructing and using composite indicators for policy makers, academics, the media and other interested parties. While there are several types of composite indicators, this Handbook is concerned with those which compare and rank country performance in areas such as industrial competitiveness, sustainable development, globalisation and innovation. The Handbook aims to contribute to a better understanding of the complexity of composite indicators and to an improvement of the techniques currently used to build them. In particular, it contains a set of technical guidelines that can help constructors of composite indicators to improve the quality of their outputs.

It has been prepared jointly by the OECD (the Statistics Directorate and the Directorate for Science, Technology and Industry) and the Applied Statistics and Econometrics Unit of the Joint Research Centre of the European Commission in Ispra, Italy. Primary authors from the JRC are Michela Nardo, Michaela Saisana, Andrea Saltelli and Stefano Tarantola. Primary authors from the OECD are Anders Hoffmann and Enrico Giovannini. Editorial assistance was provided by Candice Stevens, Gunseli Baygan and Karsten Olsen.

The research is partly funded by the European Commission, Research Directorate, under the project KEI (Knowledge Economy Indicators), Contract FP6 No. 502529. In the OECD context, the work has benefitted from a grant from the Danish government. The views expressed are those of the authors and should not be regarded as stating an official position of either the European Commission or the OECD.

Ce Manuel a pour objectif de procurer aux responsables politiques, universitaires, médias et autres parties concernées un guide sur la façon d'élaborer et d'utiliser des indicateurs composites. Si il existe plusieurs types d'indicateurs composites, ce Manuel intéresse ceux qui comparent et classent la performance d'un pays dans des domaines comme la compétitivité industrielle, le développement durable, la mondialisation et les innovations. Le Manuel a pour objectif de contribuer à une meilleure compréhension de la complexité des indicateurs composites et à une amélioration des techniques actuellement utilisées pour les élaborer. En particulier, il contient une série de lignes directrices techniques qui peuvent aider les concepteurs d'indicateurs composites à améliorer la qualité de leurs productions.

Il a été conjointement préparé par l'OCDE (la Direction des statistiques et la Direction de la science, de la technologie et de l'industrie) et la cellule des Statistiques appliquées et de l'économétrie du Centre commun de recherche (CCR) de la Commission européenne à Ispra en Italie. Les auteurs originaux du CCR sont Michela Nardo, Michaela Saisana, Andrea Saltelli et Stefano Tarantola. Les auteurs originaux de l'OCDE sont Anders Hoffmann et Enrico Giovannini. L'assistance éditoriale a été assurée par Candice Stevens, Gunseli Baygan et Karsten Olsen.

Les recherches sont partiellement financées par la Direction des recherches de la Commission européenne, pour le projet KEI (Knowledge Economy Indicators), Contrat FP6 no. 502529. Pour ce qui est de l'OCDE, le travail a bénéficié d'une subvention du gouvernement danois. Les points de vue exprimés sont ceux des auteurs et ils ne doivent pas être considérés comme l'expression d'une position officielle de la Commission européenne ou de l'OCDE.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Composite indicators (CIs) which compare country performance are increasingly recognised as a useful tool in policy analysis and public communication. They provide simple comparisons of countries that can be used to illustrate complex and sometimes elusive issues in wide ranging fields, *e.g.*, environment, economy, society or technological development. These indicators often seem easier to interpret by the general public than finding a common trend in many separate indicators and have proven useful in benchmarking country performance. However, composite indicators can send misleading policy messages if they are poorly constructed or misinterpreted. Their "big picture" results may invite users (especially policy makers) to draw simplistic analytical or policy conclusions. Instead, composite indicators must be seen as a starting point for initiating discussion and attracting public interest. Their relevance should be gauged with respect to constituencies affected by the composite index.

Pros and cons of composite indicators

In general terms, an indicator is a quantitative or a qualitative measure derived from a series of observed facts that can reveal relative positions (*e.g.*, of a country) in a given area. When evaluated at regular intervals, an indicator can point out the direction of change across different units and through time. In the context of policy analysis, indicators are useful in identifying trends and drawing attention to particular issues. They can also be helpful in setting policy priorities and in benchmarking or monitoring performance. A composite indicator is formed when individual indicators are compiled into a single index on the basis of an underlying model. The composite indicator should ideally measure multi-dimensional concepts which cannot be captured by a single indicator alone, *e.g.*, competitiveness, industrialisation, sustainability, single market integration, knowledge-based society, etc. The main pros and cons of using composite indicators are the following (**Box 1**) (Saisana and Tarantola, 2002):

Box 1. Pros and Cons of Composite Indicators

Pros	Cons
• Can summarise complex or multi-dimensional issues in view of supporting decision-makers. • Easier to interpret than trying to find a trend in many separate indicators. • Facilitate the task of ranking countries on complex issues in a benchmarking exercise. • Can assess progress of countries over time on complex issues. • Reduce the size of a set of indicators or include more information within the existing size limit. • Place issues of country performance and progress at the centre of the policy arena. • Facilitate communication with general public (i.e. citizens, media, etc.) and promote accountability.	• May send misleading policy messages if they are poorly constructed or misinterpreted. • May invite simplistic policy conclusions. • May be misused, e.g., to support a desired policy, if the construction process is not transparent and lacks sound statistical or conceptual principles. • The selection of indicators and weights could be the target of political challenge. • May disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action. • May lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored.

Composite indicators are much like mathematical or computational models. As such, their construction owes more to the craftsmanship of the modeller than to universally accepted scientific rules for encoding. As for models, the justification for a composite indicator lays in its fitness to the intended purpose and the acceptance of peers acceptance (Rosen, 1991). On the dispute whether composite indicators are good or bad as such , it has been noted:

"The aggregators believe there are two major reasons that there is value in combining indicators in some manner to produce a bottom line. They believe that such a summary statistic can indeed capture reality and is meaningful, and that stressing the bottom line is extremely useful in garnering media interest and hence the attention of policy makers. The second school, the non-aggregators, believe one should stop once an appropriate set of indicators has been created and not go the further step of producing a composite index. Their key objection to aggregation is what they see as the arbitrary nature of the weighting process by which the variables are combined." (Sharpe, 2004)

According to other commentators:

"[…] it is hard to imagine that debate on the use of composite indicators will ever be settled […] official statisticians may tend to resent composite indicators, whereby a lot of work in data collection and editing is "wasted" or "hidden" behind a single number of dubious significance. On the other hand, the temptation of stakeholders and practitioners to summarise complex and sometime elusive processes (e.g. sustainability, single market policy, etc.) into a single figure to benchmark country performance for policy consumption seems likewise irresistible." (Saisana *et al*., 2005)

Aim of the Handbook

This Handbook does not aim to resolve this debate, but only to contribute to a better understanding of the complexity of composite indicators and to an improvement of the techniques currently used to build composite indicators. In particular, it contains a set of technical guidelines that can help constructors of composite indicators to improve the quality of their outputs.

The proposal to develop a Handbook was launched at the end of a workshop on composite indicators jointly organised by the JRC and OECD in Spring 2003 which demonstrated:

- the growing interest in composite indicators from academic circles, media and policy makers;
- the existence of a wide range of methodological approaches to composite indicators; and
- the need, clearly expressed by the participants in the workshop, to have international guidelines in this domain.

Therefore, the JRC and OECD launched a project, open to other institutions, to develop the present Handbook. Key elements of the Handbook were then presented during a second workshop, held in Paris in February 2004, while the aims and the outline of the Handbook were presented at the OECD Committee on Statistics in June 2004.

The main aim of the Handbook is to provide builders of composite indicators with a set of recommendations on how to design, develop and disseminate a composite indicator. In fact, methodological issues need to be addressed transparently prior to the construction and use of composite indicators to avoid data manipulation and misrepresentation. In particular, to guide constructors and users, highlighting the technical problems and common pitfalls to be avoided, the first part of the Handbook discusses the following steps in the construction of composite indicators:

- *Theoretical framework* - A theoretical framework should be developed to provide the basis for the selection and combination of single indicators into a meaningful composite indicator under a fitness-for-purpose principle.

- *Data selection* - Indicators should be selected on the basis of their analytical soundness, measurability, country coverage, relevance to the phenomenon being measured and relationship to each other. The use of proxy variables should be considered when data are scarce.

- *Multivariate analysis* – An exploratory analysis should investigate the overall structure of the indicators, assess the suitability of the data set and explain the methodological choices, e.g., weighting, aggregation.

- *Imputation of missing data* - Consideration should be given to different approaches for imputing missing values. Extreme values should be examined as they can become unintended benchmarks.

- *Normalisation* - Indicators should be normalised to render them comparable.

- *Weighting and aggregation* – Indicators should be aggregated and weighted according to the underlying theoretical framework.

- *Robustness and sensitivity* – Analysis should be undertaken to assess the robustness of the composite indicator in terms of e.g., the mechanism for including or excluding single indicators, the normalisation scheme, the imputation of missing data and the choice of weights.

- *Links to other variables* – Attempts should be made to correlate the composite indicator with other published indicators as well as to identify linkages through regressions.

- *Visualisation* – Composite indicators can be visualised or presented in a number of different ways, which can influence their interpretation.

- *Back to the real data* - Composite indicators should be transparent and be able to be decomposed into their underlying indicators or values.

The second part of the Handbook presents a quality framework for composite indicators, where the relationships between methodologies used to construct and disseminate composite indicators and different quality dimensions are discussed. Finally, in the third part, methodologies to be used in the various steps are presented and discussed in more detail in the *Toolbox for Constructors*.

In order to help the reader in better understanding the content of the Handbook, a concrete example (the Technology Achievement Index - TAI) is used to explain the various steps in the construction of a composite indicator and highlight problems that may arise (Box 2). The TAI is a composite indicator developed by the United Nations for the Human Development Report (UN, 2001). It is composed of a relatively small number of sub-indicators, which renders it suitable for the didactic purposes of this Handbook. Moreover, the TAI is well documented by its developers and the underlying data are freely available on the Internet. For explanatory purposes, only the first 23 of the 72 original countries measured by the TAI are considered here. Further details are given in the Appendix.

The following notations are adopted throughout the Handbook:

$x^{t}_{q,c}$: raw value of sub-indicator q for country c at time t, with $q=1,\dots,Q$ and $c=1,\dots,M$

$I^{t}_{q,c}$: normalised value of sub-indicator

$w_{r,q}$: weight associated to sub-indicator q, with $r=1,\dots,R$ denoting the weighting method

CI_c^t: value of the composite indicator for country *c* at time *t*.

For reasons of clarity, the time suffix has been normally omitted and is present only in certain sections. When no time indication is present, the reader should consider that all variables have the same time dimension.

Box 2. Case study: Technology Achievement Index (TAI)

The TAI focuses on four dimensions of technological capacity (Table A.1):

- **Creation of technology**. Two sub-indicators are used to capture the level of innovation in a society: 1) the number of patents granted per capita (to reflect the current level of innovative activities) and 2) receipts from royalty and license fees from abroad per capita (to reflect the stock of successful innovations that are still useful and hence have market value).

- **Diffusion of recent innovations**. Diffusion is measured by two sub-indicators: 1) diffusion of the Internet (indispensable to participation) and 2) exports of high-and medium-technology products as a share of all exports.

- **Diffusion of old innovations**. Two sub-indicators are included: telephones and electricity. These are needed to use newer technologies and have wide-ranging applications. Both indicators are expressed as logarithms, as they are important at the earlier stages of technological advance, but not at the most advanced stages. Expressing the measure in logarithms ensures that as the level increases, it contributes less to technology achievement.

- **Human skills.** Two sub-indicators are used to reflect the human skills needed to create and absorb innovations: 1) mean years of schooling and 2) gross enrolment ratio of tertiary students enrolled in science, mathematics and engineering.

Limits of the Handbook

The literature on composite indicators is huge and almost every month new proposals on specific methodological aspects potentially relevant for the development of composite indicators are published. In this Handbook, taking into account its potential audience, we have preferred to make reference to relatively well established methodologies and procedures, avoiding the inclusion of some interesting, but still experimental, approaches. However, the Handbook should be seen as a "live" product, with successive editions, as long as new developments take place. On the other hand, this first version of the Handbook does not cover "composite leading indicators" normally used to identify cyclical movements of the economic activity. Although the OECD has a longstanding tradition and experience in this field, we have preferred to exclude them, because they are based on statistical and econometric approaches quite different from those relevant for other types of composite indicators.

I. CONSTRUCTING A COMPOSITE INDICATOR

Introduction

As already explained, the Handbook presents its recommendations following an "ideal sequence" of ten steps, from the development of a theoretical framework to the analysis of detailed data, once the indicator is built. Each step is extremely important, but the coherence of the whole process is equally important. Choices made in one step can have important implications for other steps: therefore, the composite indicator builder has not only to make the most appropriate methodological choices in each step, but also to identify if they fit well together.

Composite indicators builders have to face a relevant degree of scepticism among statisticians, economists and other groups of users. This scepticism is partially due to the lack of transparency of some existing indicators, especially as far as methodologies and basic data are concerned. To avoid these risks, the Handbook puts special emphasis on documentation and metadata. In particular, the Handbook recommends the preparation of relevant documentation at the end of each phase, both to ensure the coherence of the whole process and to prepare in advance the methodological notes that will be disseminated together with the numeric results.

This chapter provides an overview of individual steps in the construction of composite indicators. More detailed information about tools to be used in each phase is presented in Chapter III.

Step 1. Developing a theoretical framework

What is badly defined is likely to be badly measured …

A sound theoretical framework is the starting point in constructing composite indicators. The framework should clearly define the phenomenon to be measured and its sub-components and select individual indicators and weights that reflect their relative importance and the dimensions of the overall composite. Ideally, this process would be based on what is desirable to measure and not which indicators are available.

For example, gross domestic product (GDP) measures the total value of goods and services produced in a given country, where the weights are estimated based on economic theory and reflect the relative price of goods and services. The theoretical and statistical frameworks to measure GDP have been developed over the last 50 years and a revision of the 1993 System of National Accounts is currently being undertaken by the major international organisations. However, not all multi-dimensional concepts have such solid theoretical and empirical underpinnings. Composite indicators in newly emerging policy areas, *e.g.*, competitiveness, sustainable development, e-business readiness, etc., could be very subjective as the economic research in these fields is still being developed. Transparency thus is essential in constructing credible indicators. This requires:

- Defining the concept. The definition should give the reader a clear sense of what is being measured by the composite indicator. It should refer to the theoretical framework, linking various sub-groups and the underlying indicators. For example, the Growth Competitiveness Index (GCI) developed by the World Economic Forum is founded on the idea "that the process of economic growth can be analysed within three important broad categories: the macroeconomic environment, the quality of public institutions, and technology." Some complex concepts, however, are difficult to define and measure precisely or could be subject to controversy among stakeholders. Ultimately, the users of composite indicators should assess their quality and relevance.

- Determining sub-groups. Multi-dimensional concepts can be divided into several sub-groups. These sub-groups need not be (statistically) independent of each other, and existing linkages should be described theoretically or empirically to the extent possible. The Technology Achievement Index, for example, is conceptually divided into four groups of technological capacity: creation of technology, diffusion of recent innovations, diffusion of old innovations and human skills. Such a nested structure improves the users' understanding of the driving forces behind the composite indicator. It could also make it easier to determine the relative weights across different factors. This step, and also the next one, should involve experts and stakeholders as much as possible so that multiple viewpoints are acknowledged and the conceptual framework and the set of indicators gain in robustness.

- Identifying the selection criteria for the underlying indicators. The selection criteria should work as a guide for whether an indicator should be included or not in the overall composite index. It should be as precise as possible and describe the phenomenon that is being measured, i.e., input, output or process. Too often composite indicators include both input and output measures. For example, an Innovation Index could combine R&D expenditures (inputs) and the number of new products and services (outputs) in order to measure the scope of innovative activity in a given country. However, only the latter set of output indicators should be included (or expressed in terms of output per unit of input) if the index is intended to measure innovation performance.

After Step 1. the constructor should have…

- A clear understanding and definition of the multidimensional phenomenon to be measured.
- A nested structure of the various sub-groups of the phenomenon.
- A list of selection criteria for the underlying variables, *e.g.*, input, output, process.
- Clear documentation of the above.

Step 2. Selecting variables

A composite indicator is above all the sum of its parts…

The strengths and weaknesses of composite indicators largely derive from the quality of the underlying variables. Ideally, variables should be selected on the basis of their relevance, analytical soundness, timeliness, accessibility, etc. Criteria for assuring the quality of the basic data set for composite indicators are discussed in the section of this Handbook on the *Quality Framework for Composite Indicators*. While the choice of indicators must be guided by the theoretical framework for the composite, the data selection process can be quite subjective as there may be no single definitive set of indicators. The lack of relevant data also limits the constructor's ability to build sound composite indicators. Given a scarcity of internationally comparable quantitative (hard) data, composite indicators often include qualitative (soft) data from surveys or policy reviews.

Proxy measures can be used when the desired data is unavailable or when cross-country comparability is limited. For example, data on the number of employees that use computers might not be available. Instead, the number of employees who have access to computers could be used as a proxy. As in the case of soft data, caution must be given to the utilisation of proxy indicators. To the extent data permit, the accuracy of proxy measures should be checked through correlation and sensitivity analysis. The constructor should also pay close attention to whether the indicator in question is dependent on GDP or other size-related factors. To have an objective comparison across small and large countries, scaling of variables by an appropriate size measure, *e.g.*, population, income, trade volume, and populated land area, etc. is required. Finally, one has to make sure the type of the selected variables -- input, output or process indicators -- match the definition of the intended composite indicator.

The quality and accuracy of composite indicators should evolve in parallel with improvements in data collection and indicator development. The current trend towards constructing composite indicators of country performance in a range of policy areas may provide further impetus to improving data collection, identifying new data sources and enhancing the international comparability of statistics.

After Step 2. the constructor should have…

- Checked the quality of the available indicators.
- Discussed the strengths and weaknesses of each selected indicator.
- Made scale adjustments, if necessary.
- Created a summary table on data characteristics, *e.g.*, availability (across country, time), source, type (hard, soft or input, output, process)

Step 3. Multivariate analysis

Analysing the underlying structure of the data is still an art …

Over the last few decades, there has been an increase in the number of composite indicators developed by various national and international agencies. Unfortunately, individual indicators are sometimes selected in an arbitrary manner with little attention paid to the interrelationships between them. This can lead to indices which overwhelm, confuse and mislead decision-makers and the general public. Some analysts characterise this environment as ‘indicator rich but information poor’. The underlying nature of the data needs to be carefully analysed before the construction of a composite indicator. This preliminary step is helpful in assessing the suitability of the data set and will provide an understanding of the implications of the methodological choices, e.g., weighting and aggregation, during the construction phase of the composite indicator. Information can be grouped and analysed along at least two dimensions of the dataset: sub-indicators and countries.

- **Grouping information on sub-indicators**. The analyst must first decide whether the nested structure of the composite indicator is well-defined (see Step 1) and if the set of available sub-indicators is sufficient or appropriate to describe the phenomenon (see Step 2). This decision can be based on expert opinion and the statistical structure of the data set. Different analytical approaches, such as principal components analysis, can be used to explore whether the dimensions of the phenomenon are statistically well-balanced in the composite indicator. If not, a revision of the sub-indicators might be needed.

  The goal of principal components analysis (PCA) is to reveal how different variables change in relation to each other and how they are associated. This is achieved by transforming correlated variables into a new set of uncorrelated variables using a covariance matrix or its standardised form – the correlation matrix. Factor analysis (FA) is similar to PCA, however it is based on a particular statistical model. An alternative way to investigate the degree of correlation among a set of variables is to use the Cronbach coefficient alpha (c-alpha), which is the most common estimate of internal consistency of items in a model or survey. These multivariate analysis techniques are useful for gaining insight into the structure of the data set of the composite. However, it is important to avoid carrying out multivariate analysis if the sample is small compared to the number of indicators since results will not have known statistical properties.

- **Grouping information on countries**. Cluster analysis is another tool for classifying large amounts of information into manageable sets. It has been applied to a wide variety of research problems and fields from medicine to psychiatry and archaeology. Cluster analysis is also used in developing composite indicators to group information on countries based on their similarity on different sub-indicators. Cluster analysis serves as: i) a purely statistical method of aggregation of the indicators, ii) a diagnostic tool for exploring the impact of the methodological choices made during the construction phase of the composite indicator, iii) a method of disseminating information on the composite indicator without losing that on the dimensions of the sub-indicators, and iv) a method for selecting groups of countries to impute missing data with a view to decreasing the variance of the imputed values.

  When the number of variables is large or when is it believed that some of these do not contribute to identify the clustering structure in the data set, continuous and discrete models can be applied sequentially. Researchers frequently carry out a PCA and then apply a clustering algorithm on the object scores on the first few components, called "tandem analysis". However, caution is required as PCA or FA may identify dimensions that do not necessarily contribute to revealing the clustering structure in the data and may mask the taxonomic information (Table 1).

  Various alternative methods combining cluster analysis and the search for a low-dimensional representation have been proposed and focus on multidimensional scaling or unfolding analysis. Factorial k-means analysis combines k-means cluster analysis with aspects of FA and PCA. A discrete clustering model together with a continuous factorial one are fitted simultaneously to two-way data to identify the best partition of the objects, described by the best orthogonal linear combinations of the variables (factors) according to the least-squares criterion. This has a wide range of applications since it reaches a double objective: data reduction and synthesis, simultaneously in the direction of objects and variables. Originally applied to short-term macroeconomic data, factorial k-means analysis has a fast alternating least-squares algorithm that extends its application to large data sets. This methodology can be recommended as an alternative to the widely used tandem analysis.

After Step 3, the constructor should have…

- Checked the underlying structure of the data along various dimensions, *i.e.*, sub-indicators, countries.
- Applied the suitable multivariate methodology, *e.g.*, PCA, FA, cluster analysis.
- Identified sub-groups of indicators or groups of countries that are statistically “similar”.
- Analysed the structure of the data set and compared this to the theoretical framework.
- Documented the results of the multivariate analysis and the interpretation of components and factors.

Table 1. Strength and weaknesses of multivariate analysis

	Strengths	Weaknesses
Principal Components/ /Factor Analysis	Can summarise a set of sub-indicators while preserving the maximum possible proportion of the total variation in the original data set. Largest factor loadings are assigned to the sub-indicators that have the largest variation across countries, a desirable property for cross-country comparisons, as sub-indicators that are similar across countries are of little interest and cannot possibly explain differences in performance.	Correlations do not necessarily represent the real influence of the sub-indicators on the phenomenon being measured. Sensitive to modifications in the basic data: data revisions and updates, *e.g.*, new countries. Sensitive to the presence of outliers, which may introduce a spurious variability in the data. Sensitive to small-sample problems, which are particularly relevant when the focus is on a limited set of countries. Minimisation of the contribution of sub-indicators which do not move with other sub-indicators.
Cronbach Coefficient Alpha	Measures the internal consistency in the set of sub-indicators, *i.e.*, how well they describe a unidimensional construct. Thus it is useful to cluster similar objects.	Correlations do not necessarily represent the real influence of the sub-indicators on the phenomenon expressed by the composite indicator. Meaningful only when the composite indicator is computed as a 'scale' (i.e. as the sum of the sub-indicators).
Cluster Analysis	Offers a different way to group countries; gives some insight into the structure of the data set.	Purely a descriptive tool; may not be transparent if the methodological choices made during the analysis are not motivated and clearly explained.

Step 4. Imputation of missing data

The idea of imputation could be both seductive and dangerous …

Missing data often hinder the development of robust composite indicators. Data can be missing in a random or non-random fashion. The missing patterns could be:

1. *Missing completely at random* (MCAR)**.** Missing values do not depend on the variable of interest or any other observed variable in the data set. For example, the missing values in variable income would be of the MCAR type if (i) people who do not report their income have, on average, the same income as people who do report income; and if (ii) each of the other variables in the dataset would have to be the same, on average, for the people who did not report the income and the people who did report their income.

2. *Missing at random* (MAR)**.** Missing values do not depend on the variable of interest, but they are conditional on other variables in the data set. For example, the missing values in income would be MAR, if the probability of missing data on income depends on marital status but, within each category of marital status, the probability of missing income is unrelated to the value of income. Missing by design, *e.g.*, if survey question 1 is answered yes, then survey question 2 is not to be answered, are also MAR as missingness depends on the covariates.

3. *Not missing at random* (NMAR)**.** Missing values depend on the values themselves. For example, high income households are less likely to report their income.

However, there is no statistical test for NMAR and often no basis to judge whether data are missing at random or systematically, while most of the methods that impute missing values require a missing at random mechanism, *i.e.*, MCAR or a MAR. When there are reasons to assume a non-random missing pattern (NMAR), the pattern must be explicitly modelled and included in the analysis. This could be very difficult and could imply ad hoc assumptions that are likely to influence the result of the entire exercise.

There are three general methods for dealing with missing data: i) case deletion, ii) single imputation or iii) multiple imputation. The first one, also called complete case analysis, simply omits the missing records from the analysis. However, this approach ignores possible systematic differences between complete and in-complete samples and produces unbiased estimates only if deleted records are a random sub-sample of the original sample (MCAR assumption). Furthermore, standard errors will in general be larger in a reduced sample given that less information is used. As a rule of thumb, if a variable has more than 5% missing values, cases are not deleted (Little and Rubin, 2002).

The other two approaches consider the missing data as part of the analysis and try to impute values through either single imputation, *e.g.*, mean/median/mode substitution, regression imputation, hot- and cold-deck imputation, expectation-maximisation imputation, or multiple imputation, *e.g.*, Markov Chain Monte Carlo algorithm. Data imputation could lead to the minimisation of bias and the use of 'expensive to collect' data that would otherwise be discarded by case deletion. However, it can also allow data to influence the type of imputation. In the words of Dempster and Rubin (1983), "*The idea of imputation is both seductive and dangerous. It is seductive because it can lull the user into the pleasurable state of believing that the data are complete after all, and it is dangerous because it lumps together situations where the problem is sufficiently minor that it can be legitimately handled in this way and situations where standard estimators applied to real and imputed data have substantial bias.* "

The uncertainty in the imputed data should be reflected by variance estimates. This allows taking into account the effects of imputation in the course of the analysis. However, single imputation is known to underestimate the variance, because it reflects partially the imputation uncertainty. The multiple imputation method, which provides several values for each missing value, can more effectively represent the uncertainty due to imputation.

No imputation model is free of assumptions and the imputation results should hence be thoroughly checked for their statistical properties such as distributional characteristics as well as heuristically for their meaningfulness, *e.g.*, whether negative imputed values are possible.

After Step 4, the constructor should have …

- A complete data set without missing values.
- A measure of the reliability of each imputed value so as to explore the impact of imputation on the composite indicator.
- Documented and explained the selected imputation procedure and the results.

Step 5. Normalisation of data

Avoid adding up apples and oranges …

Normalisation is required prior to any data aggregation as the indicators in a data set often have different measurement units. There exist a number of normalisation methods (Table 2) (Freudenberg, 2003; Jacobs et al., 2004):

1. *Ranking* is the simplest normalisation technique. This method is not affected by outliers and allows the performance of countries to be followed over time in terms of relative positions (rankings). Country performance in absolute terms however cannot be evaluated as information on levels are lost. Some examples that use ranking include: the Information and Communications

Technology index (Fagerberg, 2001) and the Medicare study on healthcare performance across the United States (Jencks *et al*., 2003).

2. *Standardisation* (or z-scores) converts indicators to a common scale with a mean of zero and standard deviation of one. Indicators with extreme values thus have a greater effect on the composite indicator. This might be desirable if the intention is to reward exceptional behaviour. That is, if an extremely good result on a few indicators is thought to be better than a lot of average scores. This effect can be corrected in the aggregation methodology, *e.g*., by excluding the best and worst sub-indicator scores from inclusion in the index or by assigning differential weights based on the "desirability" of the sub-indicator scores.

3. *Re-scaling* normalises indicators to have an identical range (0; 1). Extreme values/or outliers however, could distort the transformed indicator. On the other hand, re-scaling could widen the range of indicators lying within a small interval increasing the effect on the composite indicator, more than they would using the z-scores transformation.

4. *Distance to a reference* measures the relative position of a given indicator vis-à-vis a reference point. This could be a target to be reached in a given time frame. For example, the Kyoto Protocol has established an 8% reduction target for CO2 emissions by 2010 for European Union members. The reference could also be an external benchmark country. For example, the United States and Japan are often used as benchmarks for the composite indicators built in the framework of the EU Lisbon agenda. Alternatively, the reference country could be the average country of the group and would be given 1, while other countries receive scores depending on their distance from the average. Hence, standardised indicators that are higher than 1 indicate countries with above-average performance. The reference country could also be the group leader where the leading country receives 1 and the others are given percentage points away from the leader. This approach, however, is based on extreme values which could be unreliable outliers.

5. *Categorical scale* assigns a score for each indicator. Categories can be numerical, such as one, two or three stars, or qualitative, such as 'fully achieved', 'partly achieved' or 'not achieved'. Often, the scores are based on the percentiles of the distribution of the indicator across countries. For example, the top 5% receive a score of 100, the units between the 85th and 95th percentiles receive 80 points, the 65th and the 85th percentiles receive 60 points, all the way to 0 points, rewarding the best performing countries and penalising the worst. Since the same percentile transformation is used for different years, any change in the definition of the indicator over time will not affect the transformed variable. However, it is difficult to follow improvements over time. Categorical scales exclude large amounts of information about the variance of the transformed indicators. Besides, when there is little variation within the original scores, the percentile bands force the categorisation on the data, irrespective of the underlying distribution. A possible solution is to adjust the percentile brackets across the individual indicators in order to obtain transformed categorical variables with almost normal distributions.

6. *Indicators* above or below the mean are transformed such that values around the mean receive 0, whereas the ones above/or below a certain threshold receive 1, and -1 respectively, e.g., the Summary Innovation Index (EC, 2001a). This normalisation method is simple and not affected by outliers. However, the arbitrariness of the threshold level and the omission of absolute level information are usually criticised. For example, if the value of a given indicator for country A is 3 times (300%) above the mean, and the value for country B is 25% above the mean, both countries would be counted as 'above average' with a threshold of 20% around the mean.

7. *Methods for cyclical indicators*. The results of business tendency surveys are usually combined into composite indicators to reduce the risk of false signals, and to better forecast cycles in economic activities (Nilsson, 2000). See, for example, the OECD composite leading indicator, and the EU economic sentiment indicators. The latter is a balance of opinions as managers of firms from different sectors and sizes are asked to express their opinion on their firm's performance. This method gives implicitly less weight to the more irregular series in the cyclical movement of the composite indicator, unless some prior ad-hoc smoothing is performed.

8. *Percentage of annual differences over consecutive years* represents the percentage growth with respect to the previous year instead of the absolute level. The transformation can be used only when the indicators are available for a number of years, e.g., Internal Market Index (EC, 2001).

The selection of a suitable method however, is not trivial and deserves special attention (Ebert and Welsh, 2004). The normalisation method should take into account the data properties, as well as the objectives of the composite indicator. Different normalisation methods will yield different results. Robustness tests might be needed to assess their impact on the outcomes.

After Step 5, the constructor should have …

- Selected the appropriate normalisation procedure(s) with reference to the theoretical framework and to the properties of the data.
- Documented and explained the selected normalisation procedure and the results.

Table 2. Normalisation methods

Method	Equation
Ranking	$I_{qc}^{t} = Rank(x_{qc}^{t})$
Standardisation (or z-scores)	$I_{qc}^{t} = \dfrac{x_{qc}^{t} - x_{qc=\bar{c}}^{t}}{\sigma_{qc=\bar{c}}^{t}}$
Re-scaling	$I_{qc}^{t} = \dfrac{x_{qc}^{t} - \min_{c}(x_{q}^{t_0})}{\max_{c}(x_{q}^{t_0}) - \min_{c}(x_{q}^{t_0})}$.
Distance to a reference country	$I_{qc}^{t} = \dfrac{x_{qc}^{t}}{x_{qc=\bar{c}}^{t_0}}$ or $I_{qc}^{t} = \dfrac{x_{qc}^{t} - x_{qc=\bar{c}}^{t_0}}{x_{qc=\bar{c}}^{t_0}}$
Categorical scales	$I_{qc}^{t} = \begin{cases} 25 & \text{if } x_{qc}^{t} \in \{p^{25th}\} \text{percentile} \\ 50 & \text{if } x_{qc}^{t} \in \{p^{50th} - p^{25th}\} \text{percentile} \\ 75 & \text{if } x_{qc}^{t} \in \{p^{75th} - p^{50th}\} \text{percentile} \\ 100 & \text{if } x_{qc}^{t} \in \{p^{100th} - p^{75th}\} \text{percentile} \end{cases}$
Indicators above or below the mean	$I_{qc}^{t} = \begin{cases} 1 & \text{if } w > (1+p) \\ 0 & \text{if } (1-p) \le w \le (1+p) \\ -1 & \text{if } w < (1+p) \end{cases}$ where $w = x_{qc}^{t} / x_{qc=\bar{c}}^{t_0}$
Cyclical indicators (OECD)	$I_{qc}^{t} = \dfrac{x_{qc}^{t} - E_t(x_{qc}^{t})}{E_t(\lvert x_{qc}^{t} - E_t(x_{qc}^{t})\rvert)}$
Balance of opinions (EC)	$I_{qc}^{t} = \dfrac{100}{N_e}\sum_{e}^{N_e} \text{sgn}_e(x_{qc}^{t} - x_{qc}^{t-1})$
Percentage of annual differences over consecutive years	$I_{qc}^{t} = \dfrac{x_{qc}^{t} - x_{qc}^{t-1}}{x_{qc}^{t}}$

Note: x_{ic}^{t} is the value of indicator for country *c* at time *t*. $\overline{c}$ is the reference country. The operator *sgn* gives the sign of the argument (i.e. +1 if the argument is positive, -1 if the argument is negative). N_e is the total number of experts surveyed.

Step 6. Weighting and aggregation

The relative importance of the indicators is a source of contention …

When used in a benchmarking framework, weights can have a significant effect on the overall composite indicator and the country rankings. A number of weighting techniques exists (**Table 3**). Some are derived from statistical models, such as factor analysis, data envelopment analysis and unobserved components models (UCM) or from participatory methods like budget allocation (BAL), analytic hierarchy processes (AHP) and conjoint analysis (CA). Unobserved components and conjoint analysis approaches are explained in the Toolbox for Constructors. No matter which method is used, weights are essentially value judgements. While some analysts might choose weights based only on statistical methods, others might reward (punish) the components that are deemed more (less) influential depending on expert opinion to better reflect the policy priorities or theoretical factors.

Most composite indicators rely on equal weighting (EW), *i.e.*, all variables are given the same weight. This could correspond to the case in which all variables are "worth" the same in the composite but also it could disguise the absence of statistical or empirical basis, *e.g.* when there is insufficient knowledge of causal relationships or a lack of consensus on the alternative. In any case, equal weighting does not mean "no weights", but implicitly implies the weights are equal. Moreover, if variables are grouped into components and those further aggregated into the composite, then applying equal weighting to the variables may imply an unequal weighting of the component (the components grouping the larger number of variables will have higher weight). This could result in an unbalanced structure of the composite index.

Weights may also be chosen to reflect the statistical quality of the data. Higher weights could be assigned to statistically reliable data with broad coverage. However, this method could be biased towards the readily available indicators, penalising the information that is statistically more problematic to identify and measure

When using equal weights, it may happen that - by combining variables with high degree of correlation - one may introduce an element of double counting into the index: if two collinear indicators are included in the composite index with a weight of w1 and w2, than the unique dimension that the two indicators measure would have weight (w1+w2) in the composite. The response has often been testing indicators for statistical correlation - for example with the Pearson correlation coefficient (Manly, 1994) - and choosing only indicators exhibiting a low degree of correlation or adjusting weights correspondingly, e.g. giving less weight to correlated indicators. Furthermore, minimizing the number of variables in the index may be desirable on other grounds such as transparency and parsimony.

Notice that there will almost always be some positive correlation between different measures of the same aggregate. Thus, a rule of thumb should be introduced to define a threshold beyond which the correlation is a symptom of double counting. On the other hand relating correlation analysis to weighting could be dangerous when motivated by apparent redundancy. For example, in the CI of e-business readiness the indicator I1 "Percentage of firms using Internet" and indicator I2 "The percentage of enterprises that have a web site" display a correlation of 0.88 in 2003: are we allowed to give less weight to the pair (I1, I2) given the high correlation or shall we consider the two indicators as measuring different aspects of innovation and communication technologies adoption and give them equal weight in constructing the composite indicator? If weights should ideally reflect the contribution of each indicator to the composite, double counting should not only be determined by statistical analysis but also by the analysis of the indicator itself *vis à vis* the rest of indicators and the phenomenon they all aim to picture.

Ideally, weights should reflect the contribution of each indicator to the overall composite. Statistical models such as principal components analysis (PCA) or factor analysis (FA) could be used to group sub-indicators. These methods account for the highest variation in the data set, using the smallest possible number of factors that reflect the underlying "statistical" dimension of the data set. Weights, however, cannot be estimated if no correlation exists between indicators. Other statistical methods, such as the "benefit of the doubt" (BOD) approach is extremely parsimonious about weighting assumptions as it lets the data decide on the weights and is sensitive to national priorities. However, weights are country specific and have a number of estimation problems.

Alternatively, participatory methods that incorporate various stakeholders -- experts, citizens and politicians -- can be used to assign weights. This approach is feasible when there is a well-defined basis for a national policy. For international comparisons, such references are often not available, or they deliver contradictory results. In the budget allocation approach, experts are given a "budget" of *N* points, to be distributed over a number of sub-indicators, "paying" more for those indicators whose importance they want to stress (Jesinghaus in Moldan and Billharz, 1997). The budget allocation is optimal for a maximum of 10-12 indicators. If too many indicators are involved, this method can give serious cognitive stress to the experts who are asked to allocate the budget. Public opinion polls have been extensively used over the years as they are easy and inexpensive to carry out (Parker, 1991).

The analytic hierarchy process (AHP) (pair wise comparison of attributes) and conjoint analysis (comparison of attributes on different levels) are also widely used techniques for multi-attribute decision making, since they enable the derivation of overall attribute (*i.e.* sub-indicator) importance based on a number of rotating attribute comparisons, as opposed to simply assigning arbitrarily given weights. The resulting weights are less sensitive to errors of judgement. However, since the AHP is based on comparisons of indicator pairs, it is applicable only to low numbers of indicators.

Aggregation methods also vary. While the linear aggregation method is useful when all sub-indicators have the same measurement unit, geometric aggregations are better suited if non-comparable and strictly positive sub-indicators are expressed in different ratio-scales. The absence of synergy or conflict across the indicators is useful in applying either linear or geometric aggregation, however difficult to achieve. Furthermore, linear aggregations reward base-indicators proportionally to the weights, while geometric aggregations reward those countries with higher scores.

In both linear and geometric aggregations, weights express trade-offs between indicators. A shortcoming in one dimension thus can be offset (compensated) by a surplus in another. This implies an inconsistency between how weights are conceived (usually they measure the importance of the associated variable) and the actual meaning when geometric or linear aggregations are used. In a linear aggregation, the compensability is constant, while with geometric aggregations compensability is lower for the composite indicators with low values. In terms of policy, if compensability is admitted (as in the case of pure economic indicators) a country with low scores on one indicator will need a much higher score on the others to improve its situation, when geometric aggregation is used. Thus in benchmarking exercises, countries with low scores prefer a linear rather than a geometric aggregation. On the other hand, the marginal utility from an increase in low absolute score would be much higher than in a high absolute score under geometric aggregation. Consequently, a country would be more interested in increasing those sectors/activities/alternatives with the lowest score in order to have the highest chance to improve its position in the ranking if the aggregation is geometric rather than linear.

If one wants to assure that weights remain a measure of importance, other aggregation methods should be used, in particular methods that do not allow compensability. Moreover if different goals are equally legitimate and important, a non-compensatory logic might be necessary. This is usually the case when highly different dimensions are aggregated in the composite, as in the case of environmental indices that include physical, social and economic data. If the analyst decides that an increase in economic performance cannot compensate a loss in social cohesion, or a worsening in environmental sustainability, then neither the linear nor the geometric aggregation is suitable. A non-compensatory multi-criteria approach (MCA) could assure non-compensability by finding a compromise between two or more legitimate goals. In its basic form, this approach does not reward outliers, as it keeps only ordinal information, *i.e.* those countries having a greater advantage (disadvantage) in sub-indicators. This method, however, could be computationally costly when the number of countries is high, as the number of permutations to calculate increases exponentially (Munda, 2005).

With regard to the time element, keeping weights unchanged across time might be justified if the researcher is willing to analyse the evolution of a certain number of variables, as in the case of the evolution of the EC internal market index from 1992 to 2002. Also in the MCA, weights do not change being associated to the intrinsic value of the indicators to explain the phenomenon. If, instead, the objective of the analysis is that of defining best practices or that of setting priorities, then weights should necessarily change over time.

The absence of an “objective” way of determining weights and aggregation methods does not necessarily lead to rejection of the validity of composite indicators, as long as the entire process is transparent. The modeller's objectives must be clearly stated at the outset, and the chosen model must be checked to see to what extent it fulfils the modeller's goal.

Table 3. Compatibility between aggregation and weighting methods

Weighting methods	Aggregation methods		
	Linear[4]	Geometric[4]	Multi-criteria
EW	Yes	Yes	Yes
PCA/FA	Yes	Yes	Yes
BOD	Yes[1]	No[2]	No[2]
UCM	Yes	No[2]	No[2]
BAL	Yes	Yes	Yes
AHP	Yes	Yes	No[3]
CA	Yes	Yes	No[3]

1 normalized with the maximin method.
2 BOD requires additive aggregation, similar arguments apply to UCM
3 At least with the multi-criteria methods requiring weights as importance coefficients.
4 With both linear and geometric aggregations weights need to trade-offs and not “importance” coefficients

After Step 6, the constructor should have …

- Selected the appropriate weighting and aggregation procedure(s) with reference to the theoretical framework.
- Considered the possibility of using multiple procedures
- Documented and explained the weighting and aggregation procedures selected.

Step 7. Robustness and sensitivity

Sensitivity analysis can be used to assess the robustness of composite indicators …

Several judgment calls have to be made when constructing composite indicators, *e.g.* on the selection of indicators, data normalisation, weights and aggregation methods, etc. The robustness of the composite indicators and the underlying policy messages may thus be contested. A combination of uncertainty and sensitivity analyses can help gauge the robustness of the composite indicator and improve transparency.

Uncertainty analysis focuses on how uncertainty in the input factors propagates through the structure of the composite indicator and affects the composite indicator values. Sensitivity analysis assesses the contribution of the individual source of uncertainty to the output variance. While uncertainty analysis is used more often than sensitivity analysis and they are almost always treated separately, the iterative use of uncertainty and sensitivity analyses during the development of a composite indicator could improve its structure (Saisana *et al*., 2005*a*; Tarantola *et al.*, 2000). Ideally, all potential sources of uncertainty should be tackled: selection of sub-indicators, data quality, normalisation, weighting, aggregation method, etc. The approach taken to assess uncertainties could include the following steps:

1. inclusion and exclusion of sub-indicators.
2. modelling data error based on the available information on variance estimation.
3. using alternative editing schemes, e.g. single or multiple imputation.
4. using alternative data normalisation schemes, such as re-scaling, standardisation, use of rankings.
5. using different weighting schemes, e.g., methods from the participatory family (budget allocation, analytic hierarchy process) and endogenous weighting (benefit of the doubt).
6. using different aggregation systems, e.g., linear, geometric mean of un-scaled variable, and multi-criteria ordering.
7. using different plausible values for the weights.

The consideration of the uncertainty inherent in the development of a composite indicator is cited in very few studies. The Human Development Index produced annually since 1990 by the United Nations Development Programme has encouraged improvement in the indicators used in its formulation. "No index can be better than the data it uses. But this is an argument for improving the data, not abandoning the index." (UN, 1992). The results of the robustness analysis are generally reported as country rankings with their related uncertainty bounds, which are due to the uncertainties at play. This would enable communicating to the user the plausible range of the composite indicator values for each country. The sensitivity analysis results are generally shown in terms of the sensitivity measure for each input source of uncertainty. These sensitivity measures represent how much the uncertainty in the composite indicator for a country would be reduced if that particular input source of uncertainty were removed. The results of a sensitivity analysis are often also shown as scatter-plots with the values of the composite indicator for a country on the vertical axis and each input source of uncertainty on the horizontal axis. Scatter-plots are useful to see patterns in the input-output relationships.

After Step 7, the constructor should have…
- Identified the sources of uncertainty in the development of the composite indicator .
- Assessed the impact of the uncertainties/assumptions on the final result.
- Conducted sensitivity analysis of the inference, *e.g.* to show what sources of uncertainty are more influential in determining the relative ranking of two entities.
- Documented and explained the sensitivity analyses and the results.Step 8. Links to other variables

Step 8. Links to other variables

Composite indicators can be linked to other variables and measures ….

Composite indicators often measure concepts that are linked to well-known and measurable phenomena, *e.g.*, productivity growth, entry of new firms. These links can be used to test the explanatory power of a composite. Simple cross-plots are often the best way to illustrate such links. An indicator measuring the environment for business start-ups, for example, could be linked to entry rates of new firms, where good performance on the composite indicator of business environment would be expected to yield higher entry rates.

For example, the Technology Achievement Index (TAI) helps to assess the position of a country relative to others concerning technology achievements. Higher technology achievement should lead to higher wealth, that is, countries with a high TAI would be expected to have high GDP per capita. Correlating TAI with GDP per capita shows this link (**Figure 1**). Most countries are close to the trend line. Only Norway and Korea are clear outliers. Norway is an outlier due to revenues from oil reserves, while

Korea has long prioritised technology development as an industrial strategy to catch-up with high-income countries.

High correlation suggests high quality of the composite indicator, although the correlation analysis should not be mistaken with causality analysis. Correlation simply indicates that the variation in the two data sets is similar. A change in the indicator does not necessarily lead to a change in the composite indicator and vice versa. Countries with high GDP might invest more in technology or more technology might lead to higher GDP. The causality remains unclear in the correlation analysis. More detailed econometric analyses can be used to determine causality, *e.g.* the Granger-causality test. However, causality tests require time series for all variables which are often not available.

The correlation can be tested with different normalisation techniques and weights. The weights for the underlying indicators can, for example, be allowed to vary between 0 and 1 for each indicator and the calculations repeated 10 000 times. The correlation analysis can then be repeated for the 10 000 indicators and the highest, median and lowest possible correlation determined. Alternatively, the correlation between the composite indicator and the measurable phenomenon can be maximised or minimised by allowing weights to vary.

It should be noted that composite indicators often include some of the indicators with which they are being correlated leading to double counting. For example, most composite indicators of sustainable development include some measure of GDP as a sub-component. In such cases, the GDP measure should be removed from the composite indicator before running any correlation.

After Step 8, the constructor should have…

- Correlated the composite indicator with related measurable phenomena.
- Tested the links with variations of the composite indicator as determined through sensitivity analysis.
- Performed econometric analysis of the links, data permitting.
- Documented and explained the correlations and the results.

Figure 1. Link between TAI and GDP per capita, 2000

Note: The correlation is significantly different from zero at the 1% level and r^2 equals 0.47. Only OECD countries are included in the correlation, as correlation with very heterogeneous groups tends to be misleading.

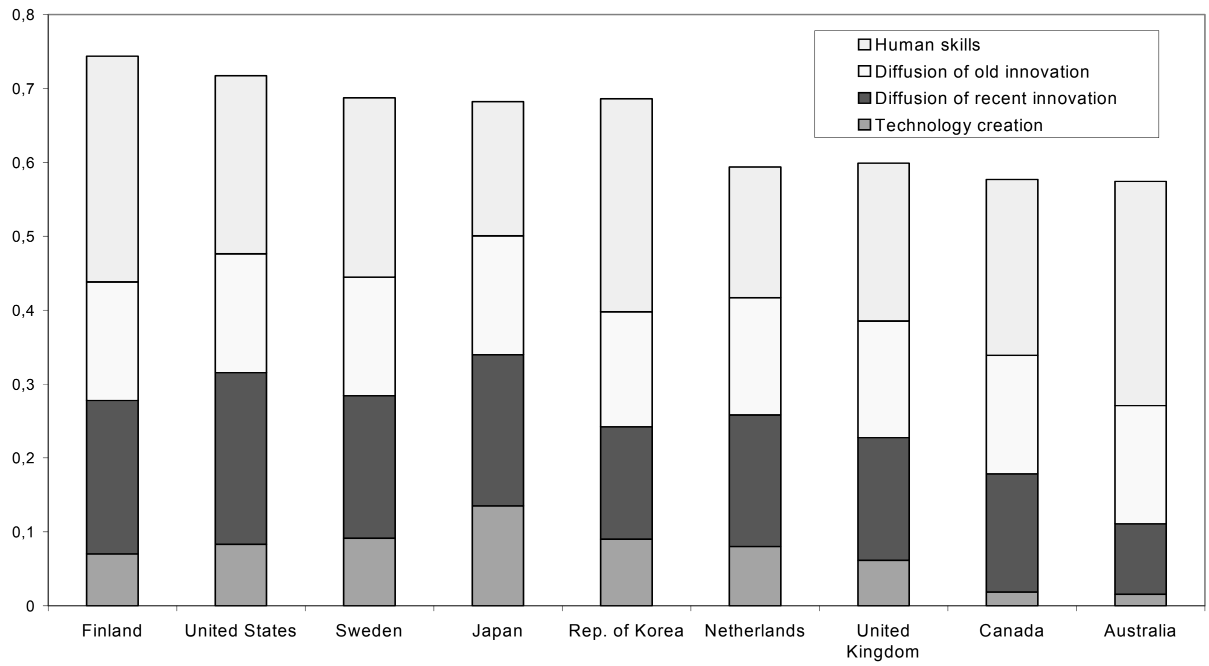
Step 9. Back to the details

De-constructing composite indicators can help extend the analysis …

Composite indicators provide a starting point for analysis. While they can be used as summary indicators to guide policy and data work, they can also be decomposed such that the contribution of sub-components and individual indicators can be identified and the analysis of country performance can be extended.

For example, the TAI index has four sub-components, which contribute differently to the aggregated composite indicator and country rankings (**Figure 2)**. This shows that a country like Finland is very strong in human skills and diffusion of recent innovations, while Japan is strong in technology creation but weaker in human skills. The decomposition of the composite indicator can thus shed light on the overall performance of a given country.

Figure 2. Example of bar chart decomposition presentation



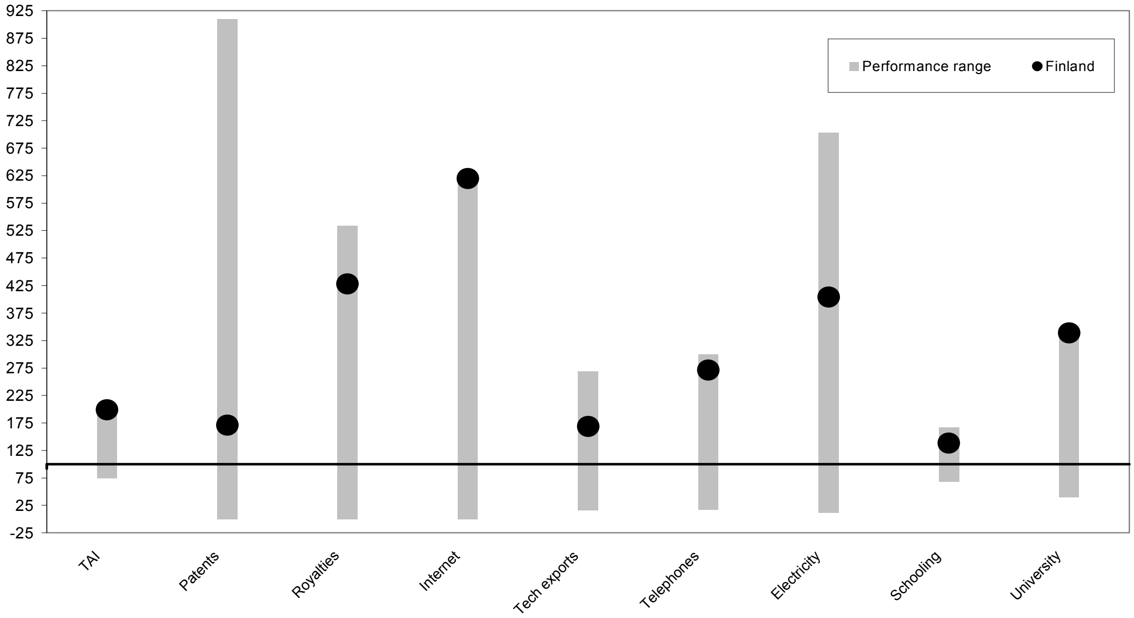
Note: Contribution of components to overall Technology Achievement Index (TAI) composite indicator. The figure is constructed by showing the standardised value of the sub-components multiplied with their individual weights. The sum of these four components equals the overall TAI index.

To profile national innovation performance**,** each sub-component of the index has been further disaggregated. The individual indicators are then used to show strengths and weaknesses. There is no optimal way of presenting individual indicators and country profiles can be presented in various ways. The following discusses three examples: 1) leaders and laggards, 2) spider diagrams and 3) traffic light presentations.

In the first example, performance on each indicator can be compared to the leader, the laggard and average performance (**Figure 3)**. Finland's top ranking is primarily based on having the highest values for the indicators relating to the Internet and universities, while the country's only weakness relates to the patent indicator.

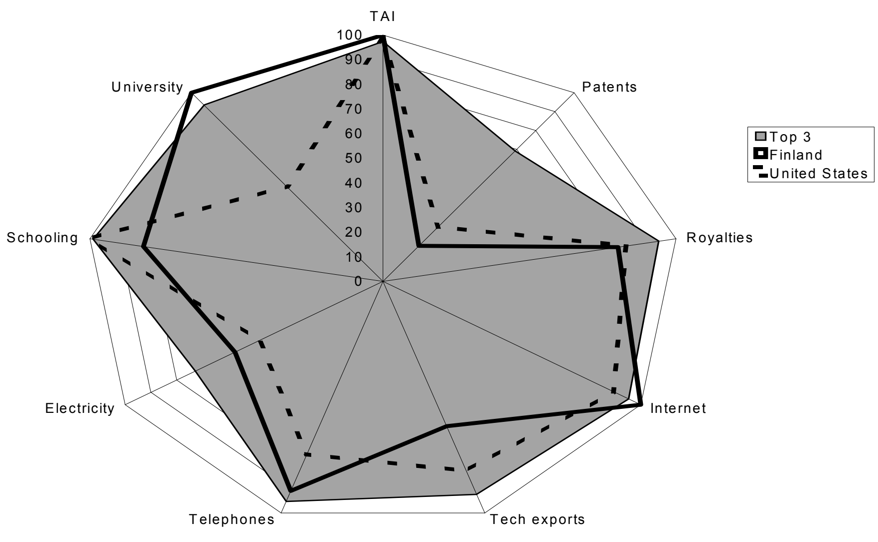
Another way of illustrating country performance is to use spider diagrams or radar charts (**Figure 4**). Here, Finland is compared to the best three countries on each indicator and another country, here the United States.

Figure 3. Example of leader/laggard decomposition presentation



Note: Technology Achievement Index (TAI). Finland (the dot) is used as an example. The figure is constructed based on the standardised indicators (distance to the mean is used). The grey area shows the range of values for that particular indicator. The average of all countries is illustrated by the 100-line.

Figure 4. Example of spider diagram decomposition presentation



Note: Technology Achievement Index (TAI). Finland is compared to the top three TAI performers and the United States. The best performing country for each indicator has the value 100 and the worst performing 0.

Finally, one can use a traffic light approach, where each indicator gets the colour green, yellow or red according to the relative performance of the country. This approach is useful when many indicators are used in the composite. For example, **Figure 5** shows that Finland has only one indicator in red (patents) and one in yellow (electricity). Japan has three in red and one in yellow.

After Step 9, the constructor should have…
• Decomposed the composite into its individual parts.
• Profiled country performance at the indicator level to reveal what is driving the aggregate results.

Documented and explained the relative

Figure 5. Example of traffic light decomposition presentation

	Index value	**Well below average (under 20)**	**Below average (20-40)**	Average (40-60)	**Above average (60-80)**	**Well above average (over 80)**
Finland						
TAI	100					
Patents	19	X				
Royalties	80					X
Internet	100					
Tech exports	63				X	
Telephones	90					X
Electricity	57			X		
Schooling	82					X
University	100					
Japan						
TAI	93					X
Patents	100					
Royalties	41			X		
Internet	24		X			
Tech exports	100					
Telephones	76				X	
Electricity	30		X			
Schooling	77				X	
University	36		X			

Note: Technology Achievement Index (TAI). Several ways exist for assigning colours. In the chosen format five colours are used but this can easily be reduced to 3. For example, green might be given to all indicators with values above the 66% quintile, yellow above 33% but below 67%.

Step 10. Presentation and dissemination

A well-designed graph can speak louder than words …

The way composite indicators are presented is not a trivial issue. Composite indicators must be able to communicate a picture to decision-makers and other end-users quickly and accurately. In particular, graphical representation of composite indicators should provide clear messages, without obscuring individual data points. On the other hand, visual presentations of composite indicators can provide signals extremely delicate from the user perspective, *e.g.,* problematic areas that require policy intervention. There are interesting ways to display and visualise composite indicators from simple tabular tools to more complicated multi-dimensional graphics and interactive software. Some examples are given below.

A tabular format is the simplest presentation where the composite indicator is presented for each country as a table of values. Usually countries are displayed in descending ranking order. Rankings can be used to track changes in country performance over time as *e.g.*, the Growth Competitiveness Index which shows the rankings of countries for two consecutive years (**Figure 6**). While tables are a comprehensive approach for displaying results, it may not be visually appealing and too detailed. However, it can be adapted to show targeted information for sets of countries grouped by geographic location, GDP, etc.

Composite indicators can be expressed via a simple bar chart (**Figure 7**). The countries are on the vertical axis and the values of the composite on the horizontal axis. The top bar indicates the average performance of all countries and enables the reader to identify how a country is performing *vis-à-vis* the average. The underlying sub-indicators can also be displayed on a bar chart. The use of colours can make the graph more visually appealing and highlight the countries performing well or not so well, growing or not growing, etc. The top bar can be thought as a target to be reached by countries.

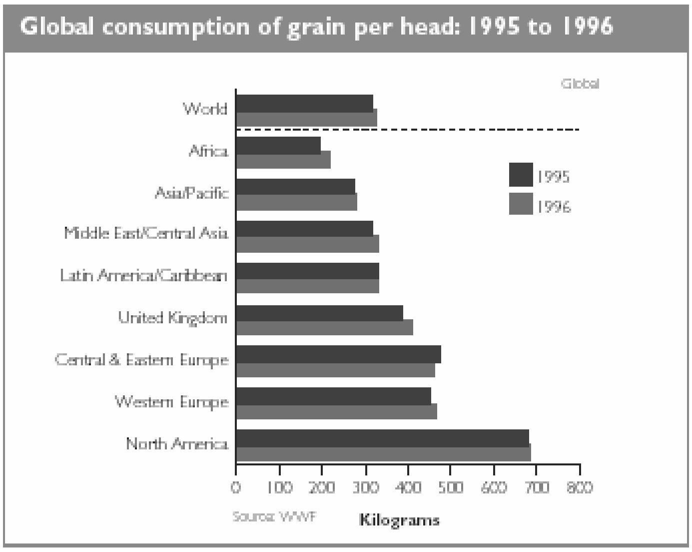
Figure 6. Example of tabular presentation of composite indicator

GROWTH COMPETITIVENESS INDEX RANKINGS

Country	Growth Competitiveness ranking 2003	Growth Competitiveness ranking 2003 among GCR 2002 countries	Growth Competitiveness ranking 2002*
Finland	1	1	1
United States	2	2	2
Sweden	3	3	3
Denmark	4	4	4
Taiwan	5	5	6
Singapore	6	6	7
Switzerland	7	7	5
Iceland	8	8	12
Norway	9	9	8
Australia	10	10	10
Japan	11	11	16
Netherlands	12	12	13
Germany	13	13	14
New Zealand	14	14	15
United Kingdom	15	15	11
Canada	16	16	9
Austria	17	17	18
Korea	18	18	25
Malta	19	—	—
Israel	20	19	17
Luxembourg	21	—	—
Estonia	22	20	27

Source: WEF, 2004.

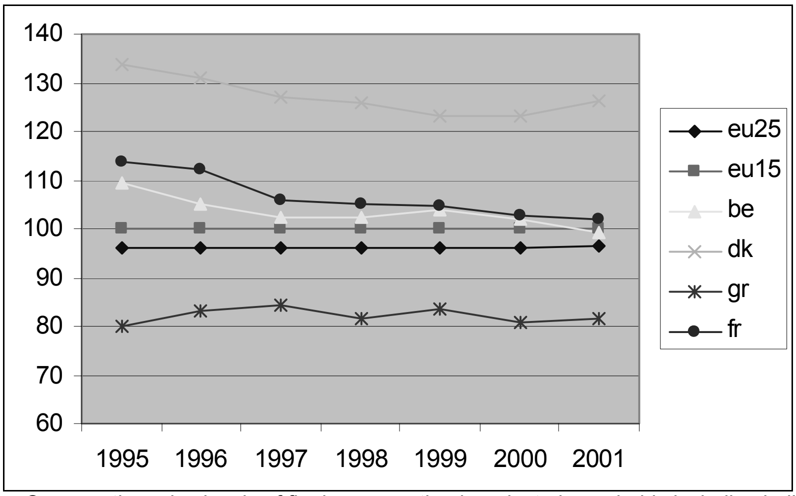
Figure 7. Example of bar chart presentation of composite indicator



Source : UK, 2004.

Line charts can be used to show performance across time. A number of lines are usually superimposed in the same chart to allow comparison between countries. Performance can be displayed *e.g.*, using a) absolute levels, b) absolute growth rates, *e.g.*, in percentage points with respect to the previous year or a number of past years, c) indexed levels and d) indexed growth rates. When indexed, the values of the indicator are linearly transformed so that their indexed value at a given year is 100. For instance, the price level index shows values such that EU15=100 at each year, with more expensive countries having values larger than 100 and less expensive countries having values lower than 100 (**Figure 8**).

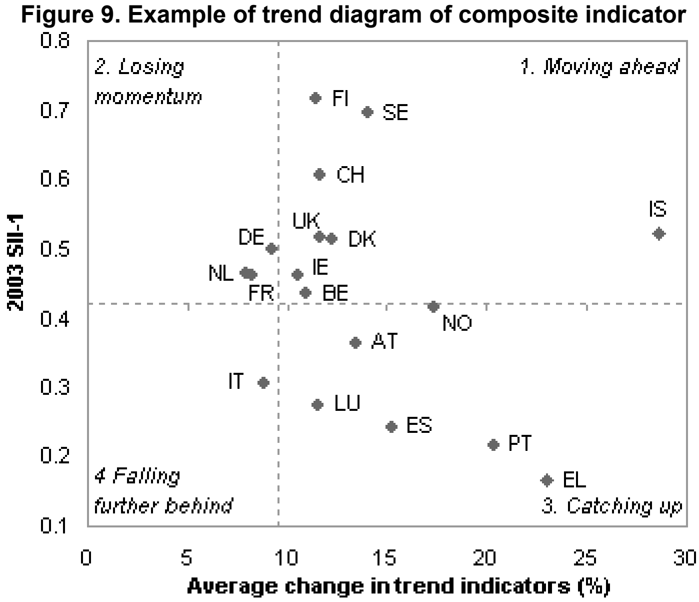
Figure 8. Example of line chart presentation of composite indicator



Note: EU price level index. Comparative price levels of final consumption by private households including indirect taxes (EU-15=100).
Source: Eurostat, 2004.

Trends in country performance as revealed through a composite indicator can be presented through trend diagrams. When a composite indicator is available for a set of countries for at least two different time points, changes or growth rates can be depicted. For example, the EU Summary Innovation Index is used to track relative performance of European countries on innovation indicators (**Figure 9**). Overall country trends are reported on the X-axis and levels are given on the Y-axis. The horizontal axis gives the EU average value and the vertical axis gives the EU trend. The two axes divide the area into four quadrants. Countries in the upper quadrant are "moving ahead", because both their value and their trend are above the EU average. Countries in the bottom left quadrant are "falling further behind" because they are below the EU average for both variables.

Figure 9. Example of trend diagram of composite indicator



Note: EU Summary Innovation Index
Source: EC, 2004.

After Step 10, the constructor should have…

- Identified a coherent set of presentational tools for the targeted audience.
- Selected the visualisation technique which communicates the most information.
- Visualised the results of the composite indicator in a clear and accurate manner.

II. QUALITY FRAMEWORK FOR COMPOSITE INDICATORS

Quality profile for composite indicators

The development of a quality framework for composite indicators is not an easy task. In fact, the overall quality of the composite indicator depends on several aspects, related both to the quality of elementary data used to build the indicator and the quality of procedures used to do it. Quality is usually defined as “fitness for use” in terms of user needs. As far as statistics are concerned, this definition is broader than has been used in the past when quality was equated with accuracy. It is now generally recognised that there are other important dimensions. Even if data are accurate, they cannot be said to be of good quality if they are produced too late to be useful, cannot be easily accessed, or appear to conflict with other data. Thus, quality is viewed as a multi-faceted concept. The most important quality characteristics depend on user perspectives, needs and priorities, which vary across groups of users.

Several organisations (*e.g.*, Statistics Canada, Statistics Sweden, Eurostat, International Monetary Fund) have been working towards the identification of various dimensions of quality for statistical products. Particularly important are the frameworks developed by the International Monetary Fund (IMF) and Eurostat. The IMF framework views quality through a prism that covers governance of statistical systems, core statistical processes and observable features of the outputs. To assess the overall quality of statistics produced by its member countries, the IMF has developed the “Data Quality Assurance Framework (DQAF)”, which addresses a broad range of questions that are captured through the i) prerequisites of quality and ii) five quality dimensions.

With regard to the prerequisites of quality, the DQAF assesses how the quality of statistics is affected by the legal and institutional environment and available resources and whether there exists quality awareness in managing statistical activities. Therefore, an evaluation of the way in which a national statistical office (or system) performs its task is carried out through a detailed questionnaire to identify the degree of scientific independence of statistical agencies, the autonomy given to statistical agencies, etc.

The five quality dimensions used by the IMF are the following:

1. *Assurance of integrity*: What are the features that support firm adherence to objectivity in the production of statistics, so as to maintain users’ confidence?

2. *Methodological soundness*: How do the current practices relate to the internationally agreed methodological practices for specific statistical activities?

3. *Accuracy and reliability*: Are the source data, statistical techniques, etc. adequate to portray the reality to be captured?

4. *Serviceability*: How are users’ needs met in terms of timeliness of the statistical products, their frequency, consistency, and their revision cycle?

5. *Accessibility*: Are effective data and metadata easily available to data users and is there assistance to users?

The Eurostat framework focuses on statistical outputs as viewed by users and works its way back to the underlying processes only where the outputs do not yield a direct measurement. It is based on seven dimensions, which try to answer to the following questions:

1. *Relevance*: are the data what the user expects?

2. *Accuracy*: are the figures reliable?

3. *Comparability*: are the data in all necessary respects comparable across countries?

4. *Completeness:* are domains for which statistics are available reflecting the needs expressed by users?

5. *Coherence*: are the data coherent with other data?

6. *Timeliness and punctuality*: does the user receive the data in time and according to pre-established dates?

7. *Accessibility and clarity*: is the figure accessible and understandable?

Given the institutional set-up of the European Statistical System, the main aim of the Eurostat quality approach is to ensure that certain standards are met in various aspects of statistical production processes carried out by national statistical agencies and by Eurostat itself. In addition, it largely aims to use quantifiable measures, such as measurement errors or days (or months) of publication delay after the reference period.

There are several areas of commonalities between the two approaches, but, notwithstanding the effort made over the last few years to further harmonise them, they are quite different in scope. For example, the IMF approach focuses on process-oriented indicators, is mainly based on qualitative assessments and was designed with national sources in mind, while the Eurostat approach focuses on output-oriented indicators, aims to provide, to the extent possible, quantitative measures and can be applied both to national and European data.

More recently, the OECD developed and published the first version of its “Quality Framework and Guidelines for OECD Statistics” (OECD, 2003). It relies heavily on the results achieved by the international statistical community, adapting them to the OECD context. In fact, for an international organisation, the quality of statistics disseminated depends on two aspects: i) the quality of national statistics received, and ii) the quality of internal processes for collection, processing, analysis and dissemination of data and metadata. From this point of view, there are some similarities between what the OECD has done in the development of its own quality framework and the characteristics of composite indicators, whose overall quality depends on two aspects: i) the quality of basic data, and ii) the quality of procedures used to build and disseminate the composite indicator.

Both elements are equally important: the application of the most advanced approaches to the development of composite indicators based on inaccurate or incoherent data would not produce high quality results. Similarly, a composite indicator which combines very good basic data but uses poor procedures would produce unreliable and unstable results. Finally, composite indicators disseminated without appropriate metadata could easily be misinterpreted. Therefore, the quality framework for composite indicators must consider all these aspects. In the following section, each are considered separately.

Quality dimensions for basic data

The selection of basic data should maximise the overall quality of the final result. In particular, in selecting these data the following dimensions are to be considered:

Relevance

The relevance of data is a qualitative assessment of the value contributed by these data. Value is characterised by the degree to which the data serves to address the purposes for which they are sought by users. It depends upon both the coverage of the required topics and the use of appropriate concepts.

> In the context of composite indicators, relevance has to be evaluated considering the overall purpose of the indicator. A careful evaluation and selection of basic data have to be carried out to ensure that the right range of domains is covered in a balanced way. Given the actual availability of data, "proxy" series are often used, but in this case some evidence about their relationships with "target" series should be produced whenever possible.

Accuracy

The accuracy of basic data is the degree to which they correctly estimate or describe the quantities or characteristics that they are designed to measure. Accuracy refers to the closeness between the values provided and the (unknown) true values. Accuracy has many attributes, and in practical terms there is no single aggregate or overall measure of it. Of necessity, these attributes are typically measured or described in terms of the error, or the potential significance of error, introduced through individual major sources of error.

In the case of sample survey-based estimates, the major sources of error include coverage, sampling, non-response, response, processing, and problems in dissemination. For derived estimates, such as for national accounts or balance of payments, sources of error arise from the surveys and censuses that provide source data; from the fact that source data do not fully meet the requirements of the accounts in terms of coverage, timing, and valuation and that the techniques used to compensate can only partially succeed; from seasonal adjustment; and from separation of price and quantity in the preparation of volume measures.

An aspect of accuracy is the closeness of the initially released value(s) to the subsequent value(s) of estimates. In light of the policy and media attention given to first estimates, a key point of interest is how close a preliminary value is to subsequent estimates. In this context it useful to consider the sources of revision, which include (1) replacement of preliminary source data with later data, (2) replacement of judgmental projections with source data, (3) changes in definitions or estimating procedures, and (4) updating of the base year for constant-price estimates. Smaller and fewer revisions is an aim; however, the absence of revisions does not necessarily mean that the data are accurate.

> In the context of composite indicators, accuracy of basic data is extremely important. Here the issue of credibility of the source becomes crucial. The credibility of data products refers to confidence that users place in those products based simply on their image of the data producer, i.e., the brand image. One important aspect is trust in the objectivity of the data. This implies that the data are perceived to be produced professionally in accordance with appropriate statistical standards and policies and that practices are transparent (for example, data are not manipulated, nor their release timed in response to political pressure). Other things equal, data produced by "official sources" (*e.g.* national statistical offices or other public bodies working under national statistical regulations or codes of conduct) should be preferred to other sources

Timeliness

The timeliness of data products reflects the length of time between their availability and the event or phenomenon they describe, but considered in the context of the time period that permits the information to be of value and still acted upon. The concept applies equally to short-term or structural data; the only difference is the timeframe. Closely related to the dimension of timeliness, the punctuality of data products is also very important, both for national and international data providers. Punctuality implies the existence of a publication schedule and reflects the degree to which data are released in accordance with it.

In the context of composite indicators, timeliness is especially important to minimise the need for estimating missing data and for revisions of previously published data. As individual basic data sources establish their optimal trade-off between accuracy and timeliness taking into account institutional, organisational and resource constraints, often data covering different domains are released at different points of time. Therefore, special attention must be paid to the overall coherence of vintages of data used to build composite indicators (see also coherence).

Accessibility

The accessibility of data products reflects how readily the data can be located and accessed from original sources. The range of different users leads to such considerations as multiple dissemination formats and selective presentation of metadata. Thus, accessibility includes the suitability of the form in which the data are available, the media of dissemination, and the availability of metadata and user support services. It also includes the affordability of the data to users in relation to its value to them and whether the user has a reasonable opportunity to know that the data are available and how to access them.

In the context of composite indicators, accessibility of basic data can affect the overall cost of production and updating of the indicator over time. It can also influence the credibility of the composite indicator if poor accessibility of basic data makes it difficult for third parties to replicate the results of the composite indicators. In this respect, given improvements in electronic access to databases released by various sources, the issue of coherence across data sets can become relevant. Therefore, the selection of the source should not always give preference to the most accessible source, but also look at other quality dimensions.

Interpretability

The interpretability of data products reflects the ease with which the user may understand and properly use and analyse the data. The adequacy of the definitions of concepts, target populations, variables and terminology underlying the data, and information describing the limitations of the data, if any, largely determines the degree of interpretability. The range of different users leads to such considerations as metadata presentation in layers of increasing detail. Definitional and procedural metadata assist in interpretability: thus, the coherence of these metadata is an aspect of interpretability.

In the context of composite indicators, the wide range of data used to build them and the difficulties due to the aggregation procedure require the full interpretability of basic data. The availability of definitions and classifications used to produce basic data is essential to assess the comparability of data over time and across countries (see coherence): for example, series breaks need to be assessed when composite indicators are built to compare performances over time. Therefore, the availability of adequate metadata is an important element to assess the overall quality of basic data.

Coherence

The coherence of data products reflects the degree to which they are logically connected and mutually consistent. Coherence implies that the same term should not be used without explanation for different concepts or data items; that different terms should not be used without explanation for the same concept or data item; and that variations in methodology that might affect data values should not be made without explanation. Coherence in its loosest sense implies the data are "at least reconcilable". For example, if two data series purporting to cover the same phenomena differ, the differences in time of recording, valuation, and coverage should be identified so that the series can be reconciled.

In the context of composite indicators, two aspects of coherence are especially important: coherence over time and across countries. *Coherence over time* implies that the data are based on common concepts, definitions, and methodology over time, or that any differences are explained and can be allowed for. Incoherence over time refers to breaks in a series resulting from changes in concepts, definitions, or methodology. *Coherence across countries* implies that from country to country the data are based on common concepts, definitions, classifications and methodology, or that any differences are explained and can be allowed for.

Quality dimensions for procedures to build and disseminate composite indicators

Each phase of the composite indicator building process is important and has to be carried out with quality concerns in mind. For example, the design of the theoretical framework can affect the relevance of the indicator; the multivariate analysis is important to increase its reliability; the imputation of missing data, as well as the normalisation and the aggregation, can affect its accuracy, etc. In the following matrix, the most important links between each phase of the building process and quality dimensions are identified, using the seven dimensions of the OECD Quality Framework (**Table** 4).

The proper definition of the theoretical framework affects the relevance of the composite indicator, but also its credibility and interpretability. The relevance of a composite indicator is usually evaluated taking into account analytical and policy needs, but also its theoretical foundation. From this point of view, several composite indicators are quite weak and such weakness is often quoted to criticise the overall idea of composite indicators.

The quality of basic data chosen to build the composite indicator strongly affects its accuracy and credibility. Also timeliness can be largely influenced by the choice of appropriate data. The use of multivariate analysis to identify the data structure can increase both the accuracy and the interpretability of final results. This step is also very important to identify redundancies among selected phenomena and evaluate possible gaps in basic data.

The imputation of missing data affects the accuracy of the composite indicator and its credibility. Furthermore, too much use of imputation techniques can undermine the overall quality of the indicator and its relevance, even if it can improve the dimension of timeliness. The normalisation phase is crucial both for the accuracy and the coherence of final results. A inappropriate normalisation procedure can bring about unreliable or biased results. On the other hand, the interpretability of the composite indicator heavily relies on the correctness of the approach followed in the normalisation phase.

One of the key issues in the construction of composite indicators is the choice of the weighting and aggregation model. Almost all quality dimensions are affected by this choice, especially accuracy, coherence and interpretability. This is also one of the most criticised characteristics of composite indicators: therefore, the indicator builder has to pay special attention to avoid internal contradictions and mistakes when dealing with weighting and aggregating individual indicators.

To minimise the risks of producing meaningless composite indicators, sensitivity and robustness analyses are needed. Analysis of this type can improve the accuracy, credibility and interpretability of the final results. Given public and media interest in country rankings, sensitivity checks can help distinguish significant and insignificant differences, minimising the risk of misinterpretation and misuse.

The comparison between the composite indicator and other well known and "classical" measures of relevant phenomena can be very useful to evaluate the capacity of the former to produce meaningful and relevant results. Therefore, relevance and interpretability of the results can be strongly reinforced by such comparison. In addition, the credibility of the indicator can benefit by its capacity to produce results which are highly correlated with the reference data.

The presentation of composite indicators and their visualisation affects both relevance and interpretability of the results. Given the complexity of composite indicators, the general public (media, citizens, etc.), as well as policy makers, will not generally read methodological notes and "caveats". Therefore, their comprehension of the results will be largely based on the "messages" given through summary tables or charts.

As highlighted in this Handbook, composite indicators provide a starting point for analysis, which has to be deepened going back to the detail. Therefore, this analytical phase can affect the relevance of the indicator and also its interpretability. Moreover, if the way in which the indicator is built or disseminated does not allow users and analysts to go into the details, the overall credibility of the exercise can be affected.

Finally, the dissemination phase is crucial to assure the relevance of the indicator, its credibility, accessibility and interpretability. Too often statisticians do not pay enough attention to this fundamental phase, thus limiting the audience for their products and their overall impact. The OECD has recently developed a Handbook for data and metadata presentation which contains useful practices to improve the dissemination of statistical products.

Table 4. Quality dimensions of composite indicators

CONSTRUCTION PHASE	**QUALITY DIMENSIONS**						
	Relevance	*Accuracy*	*Credibility*	*Timeliness*	*Accessibility*	*Interpretability*	*Coherence*
Theoretical framework	X		X			X	
Data selection		X	X	X			
Multivariate analysis		X				X	X
Imputation of missing data	X	X	X	X			
Normalisation		X				X	X
Weighting and aggregation	X	X	X			X	X
Robustness and sensitivity		X	X			X	
Links to other variables	X		X			X	X
Visualisation	X					X	
Back to the data	X		X			X	
Dissemination	X		X		X	X	

1. For explanatory purposes, only the first 23 of the 72 original countries measured by the TAI are considered here. Further details are given in the Appendix.

III. TOOLBOX FOR CONSTRUCTORS

A number of statistical methods are discussed here in detail to provide constructors the necessary tools for building sound composite indicators, focusing on the practical implementation of the steps previously outlined. The need for multivariate analysis prior to the aggregation of the individual indicators is stressed. Also discussed are the problem of missing data and the techniques used to standardise indicators of a very different nature into a common unit. Different methodologies for weighting and aggregating indicators into a composite are explored as well as the need to test the robustness of the composite using uncertainty and sensitivity analysis. The example of the Technology Achievement Index (TAI) (see Appendix) is used as a baseline case to illustrate differences across different methods and to highlight potential pitfalls.

MULTIVARIATE ANALYSIS

Multivariate data analysis techniques which have found use inthe construction or analysis of composite indicators are described in this section.

Principal components analysis

The objective[1] is to explain the variance of the observed data through a few linear combinations of the original data. Even though there are Q variables, $x_1, x_2, \ldots x_Q$, much of the data's variation can often be accounted for by a small number of variables – principal components, or linear relations of the original data, $Z_1, Z_2, \ldots Z_Q$ that are uncorrelated. At this point there are still Q principal components, *i.e.*, as many as there are variables. The next step is to select the first, say P<Q principal components that preserve a "high" amount of the cumulative variance of the original data.

$$
\begin{aligned}
Z_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1Q}x_Q \\
Z_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2Q}x_Q \\
&\ldots \\
Z_Q &= a_{Q1}x_1 + a_{Q2}x_2 + \ldots + a_{QQ}x_Q
\end{aligned}
\qquad \textbf{(1)}
$$

The lack of correlation in the principal components is a useful property. It indicates that the principal components are measuring different "statistical dimensions" in the data. When the objective of the analysis is to present a huge dataset using a few variables, some degree of economy can be achieved by applying Principal Components Analysis (PCA), if the variation in the Q original x variables can be accounted for by a small number of Z variables. It must be stressed that PCA cannot always reduce a large number of original variables to a small number of transformed variables. Indeed, if the original variables are uncorrelated, then the analysis is of no value. On the other hand, a significant reduction is obtained when the original variables are highly correlated--positively or negatively.

The weights a_{ij} (also called component or factor loadings) applied to the variables x_j in Equation (1) are chosen so that the principal components Z_i satisfy the following conditions:

(i) they are uncorrelated (orthogonal),

(ii) the first principal component accounts for the maximum possible proportion of the variance of the set of x s, the second principal component accounts for the maximum of the remaining variance and so on until the last of the principal component absorbs all the remaining variance not accounted for by the preceding components, and [2]

$$\alpha_{i1}^2 + \alpha_{i2}^2 + \ldots + \alpha_{iQ}^2 = 1, i = 1,2,\ldots,Q$$

PCA involves finding the *eigenvalues* λ_j , $j=1,\ldots,Q$, of the sample covariance matrix *CM*,

$$CM = \begin{bmatrix} cm_{11}\ cm_{12}\ \ldots cm_{1Q} \\ cm_{21}\ cm_{22}\ \ldots cm_{2Q} \\ \ldots \\ cm_{Q1}\ cm_{Q2}\ \ldots cm_{QQ} \end{bmatrix} \quad \textbf{(2)}$$

where, the diagonal element cm_{ii} is the variance of x_i and cm_{ij} is the covariance of variables x_i and x_j. The eigenvalues of the matrix *CM* are the variances of the principal components and can be found by solving the characteristic equation. $|CM - \lambda I| = 0$ where *I* is the identity matrix with the same order as CM, and λ is the vector of eigenvalues. This is possible, however, only if Q is small. If there are too many variables solving for λ is non-trivial and other methods exist (see i.e. Gentle, Härdle & Mori, 2004; and Golub & van der Vorst, 2000)[3]. There are Q eigenvalues, some of which may be negligible. Negative eigenvalues are not possible for a covariance matrix. An important property of the eigenvalues is that they add up to the sum of the diagonal elements of CM. That is, the sum of the variances of the principal components is equal to the sum of the variances of the original variables:

$$\lambda_1 + \lambda_2 + \ldots + \lambda_Q = cm_{11} + cm_{22} + \ldots + cm_{QQ} \quad (3)$$

In order to avoid one variable having an undue influence on the principal components, it is common to standardise the variables --x s-- to have zero means and unit variances at the start of the analysis. The co-variance matrix CM then takes the form of the correlation matrix (**Table 5**). For the TAI example, the highest correlation is found between the sub-indicators ELECTRICITY & INTERNET with a coefficient of 0.84.

Table 5. Correlation matrix for TAI sub-indicators

	PATENTS	ROYALTIES	INTERNET	EXPORTS	TELEPHONE	ELECTRICITY	SCHOOLING	ENROLMENT
PATENTS	1.00	0.13	-0.09	**0.45**	0.28	0.03	0.22	0.08
ROYALTIES		1.00	**0.46**	0.25	**0.56**	0.32	0.30	0.06
INTERNET			1.00	**-0.45**	**0.56**	**0.84**	**0.63**	0.27
EXPORTS				1.00	0.00	-0.36	-0.35	-0.03
TELEPHONE					1.00	**0.64**	0.30	0.33
ELECTRICITY						1.00	**0.65**	0.26
SCHOOLING							1.00	0.08
ENROLMENT								1.00

Note: n=23. Marked correlations are statistically significant at $p < 0.05$.

Table 6 gives the eigenvalues of the correlation matrix of the eight sub-indicators (standardised values) that compose TAI. Note that the sum of the eigenvalues is equal to the number of sub-indicators ($Q = 8$).

Figure 0a is a graphical presentation of the eigenvalues in descending order. Given that the correlation matrix rather than the covariance matrix is used in the PCA, all 8 sub-indicators are assigned equal weights in forming the principal components (Chatfield and Collins, 1980). The first Principal Component explains the maximum variance in all the sub-indicators – eigenvalue of 3.3. The second principal component explains the maximum amount of the remaining variance – a variance of 1.7. The third and fourth principal components have an eigenvalue close to 1. The last four principal components explain the remaining 12.8% of the variance in the dataset.

Table 6. Eigenvalues of TAI sub-indicators

	Eigenvalue	% of variance	Cumulative %
1	3.3	41.9	41.9
2	1.7	21.8	63.7
3	1.0	12.3	76.0
4	0.9	11.1	87.2
5	0.5	6.0	93.2
6	0.3	3.7	96.9
7	0.2	2.2	99.1
8	0.1	0.9	100.0

Note: Extraction method: Principal Components Analysis, n=23.

A drawback of the conventional PCA is that it does not allow for inference on the properties of the general population. Traditionally, drawing such inferences requires certain distributional assumptions to be made regarding the population characteristics, which the PCA techniques are not based upon There are several assumptions made in the application of PCA/FA which are discussed in **Box 3**. These assumptions are mentioned in almost all textbooks, yet they are often neglected when composite indicators are developed.

Furthermore, in a traditional PCA framework, there is no estimation of the statistical precision of the results, which is essential for relatively small sample sizes, as in the present case of the TAI example. Therefore, the bootstrap method has been utilised in conjunction with PCA to make inferences about the population (Efron and Tibshirani, 1991, 1993). Bootstrap refers to the process of randomly re-sampling the original data set to generate new data sets. Estimates of the relevant statistics are made for each bootstrap sample. A very large number of bootstrap samples will give satisfactory results but the computation may be cumbersome. Various values have been suggested, ranging from 25 (Efron and Tibshirani, 1991) to as high as 1000 (Efron, 1987; Mehlman *et al*., 1995).

Box 3. Assumptions in Principal Components Analysis and Factor Analysis

Enough number of cases. The question of how many cases (or countries) are necessary to do PCA/FA has no scientific answer and methodologists' opinions differ. Alternative arbitrary rules of thumb in descending order of popularity include those below.

Rule of 10. There should be at least 10 cases for each variable.

3:1 ratio. The cases-to-variables ratio should be no lower than 3 (Grossman et al. 1991).

5:1 ratio. The cases-to-variables ratio should be no lower than 5 (Bryant and Yarnold, 1995; Nunnaly 1978, Gorsuch 1983).

Rule of 100: The number of cases should be the larger between (5 × number of variables), and 100. (Hatcher, 1994).

Rule of 150: Hutcheson and Sofroniou (1999) recommend at least 150 - 300 cases, more toward 150 when there are a few highly correlated variables.

Rule of 200. There should be at least 200 cases, regardless of the cases-to-variables ratio (Gorsuch, 1983).

Significance rule. There should be 51 more cases than the number of variables, to support chi-square testing (Lawley and Maxwell, 1971)

These rules are not mutually exclusive. Bryant and Yarnold (1995), for instance, endorse both the cases-to-variables ratio and the Rule of 200. In the TAI example, there are 23:8 cases-to-variables, therefore the first and the second rule are satisfied.

No bias in selecting sub-indicators. The exclusion of relevant sub-indicators and the inclusion of irrelevant sub-indicators in the correlation matrix being factored will affect, often substantially, the factors which are uncovered. Although social scientists may be attracted to factor analysis as a way of exploring data whose structure is unknown, knowing the factorial structure in advance helps select the sub-indicators to be included and yields the best analysis of factors. This dilemma creates a chicken-and-egg problem. Note this is not just a matter of including all relevant sub-indicators. Also, if one deletes sub-indicators arbitrarily in order to have a "cleaner" factorial solution, erroneous conclusions about the factor structure will result (Kim and Mueller, 1978a).

No outliers. As with most techniques, the presence of outliers can affect interpretations arising from PCA/FA. One may use Mahalanobis distance to identify cases, which are multivariate outliers and remove them prior to the analysis. Alternatively, one can also create a dummy variable set to 1 for cases with high Mahalanobis distance, then regress this dummy on all other variables. If this regression is non-significant (or simply has a low R-squared for large samples) then the outliers are judged to be at random and there is less danger in retaining them. The ratio of the regression coefficients indicates which variables are most associated with the outlier cases.

Assumption of interval data. Kim and Mueller (1978b) note that ordinal data may be used if it is thought that the assignment of ordinal categories to the data does not seriously distort the underlying metric scaling. Likewise, the use of dichotomous data is allowed, if the underlying metric correlation between the variables are thought to be moderate (.7) or lower. The result of using ordinal data is that the factors may be much harder to interpret. Note that categorical variables with similar splits will necessarily tend to correlate with each other, regardless of their content (see Gorsuch, 1983). This is particularly apt to occur when dichotomies are used. The correlation will reflect similarity of "difficulty" for items in a testing context; hence such correlated variables are called *difficulty factors*. The researcher should examine the factor loadings of categorical variables with care to assess whether common loading reflects a difficulty factor or substantive correlation.

Linearity. Principal components factor analysis (PFA), which is the most common variant of FA, is a linear procedure. Of course, as with multiple linear regression, nonlinear transformation of selected variables may be a pre-processing step, but this is not common. The smaller the sample size, the more important it is to screen data for linearity.

Multivariate normality of data is required for related significance tests. PCA and PFA have no distributional assumptions. Note, however, that a variant of factor analysis, maximum likelihood factor analysis, does assume multivariate normality. The smaller the sample size, the more important it is to screen data for normality. Moreover, as factor analysis is based on correlation (or sometimes covariance), both correlation and covariance will be attenuated when variables come from different underlying distributions (ex., a normal vs. a bimodal variable will correlate less than 1.0 even when both series are perfectly co-ordered).

Underlying dimensions shared by clusters of sub-indicators are assumed. If this assumption is not met, the "garbage in, garbage out" principle applies. Factor analysis cannot create valid dimensions (factors) if none exist in the input data. In such cases, factors generated by the factor analysis algorithm will not be comprehensible. Likewise, the inclusion of multiple definitionally-similar sub-indicators representing essentially the same data will lead to tautological results.

Strong intercorrelations are not mathematically required, but applying factor analysis to a correlation matrix with only low intercorrelations will require for solution nearly as many factors as there are original variables, thereby defeating the data reduction purposes of factor analysis. On the other hand, too high inter-correlations may indicate a multi-collinearity problem and collinear terms should be combined or otherwise eliminated prior to factor analysis.

(a) The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is a statistics for comparing the magnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients. The concept is that the partial correlations should not be very large if one is to expect distinct factors to emerge from factor analysis (Hutcheson and Sofroniou, 1999). A KMO statistic is computed for each individual sub-indicator, and their sum is the KMO overall statistic. KMO varies from 0 to 1.0. A KMO overall should be .60 or higher to proceed with factor analysis (Kaiser and Rice, 1974), though realistically it should exceed 0.80 if the results of the principal components analysis are to be reliable. If not, it is recommended to drop the sub-indicators with the lowest individual KMO statistic values, until KMO overall rises above .60.

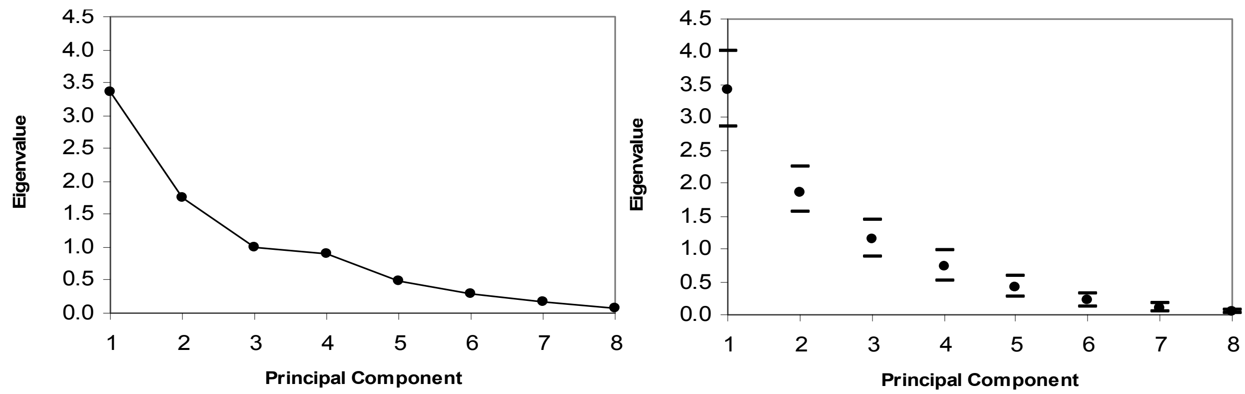
(b) Variance-inflation factor (VIF) is simply the reciprocal of tolerance. A VIF value greater than 4.0 is an arbitrary but common cut-off criterion for suggesting that there is a multi-collinearity problem. Some researchers use the more lenient cutoff VIF value of 5.0.

(c) The Bartlett's test of sphericity is used to test the null hypothesis that the sub-indicators in a correlation matrix are uncorrelated, that is to say that the correlation matrix is an identity matrix. The statistic is based on a chi-squared transformation of the determinant of the correlation matrix. However, as Bartlett's test is highly sensitive to sample size (Knapp and Swoyer 1967), Tabachnick and Fidell (1989) suggest implementing it with the KMO measure.

An important issue however, is whether the TAI dataset for the 23 countries can be viewed as a 'random' sample of the entire population, as required by the bootstrap procedures (Efron 1987; Efron and Tibshirani 1993). Several points can be made regarding the issues of randomness and representativeness of the data. First, it is often difficult to obtain complete information for a dataset in the social sciences, as controlled experiments are not always possible, unlike in natural sciences.. As Efron and Tibshirani (1993) state: 'in practice the selection process is seldom this neat […], but the conceptual framework of random sampling is still useful for understanding statistical inferences.' Second, the countries included in the restricted set show no apparent pattern as to whether or not they are predominately developed or developing countries. In addition, the countries of varying sizes span all the major continents of the world, ensuring a wide representation of the global state of technological development. Consequently, the restricted set could be considered as representative of the total population. A third point on the data quality is that a certain amount of measurement error is likely to exist. While such measurement error can only be controlled at the data collection stage, rather than at the analytical stage, it is argued that the data represent the best estimates currently available (UN, 2001).

Figure 10b demonstrates graphically the relationship between the eigenvalues from the deterministic PCA, their bootstrapped confidence intervals (5^{th} and 95^{th} percentiles) and the ranked principal components. These confidence intervals allow one to generalise the conclusions concerning the small set of the sub-indicators (23 countries) to the entire population (e.g. of 72 countries or even more general), rather than confining the conclusions only to the sample set being analysed. Bootstrapping has been performed for 1000 sample sets of size 23 (random sampling with replacement). It is shown that the values of the eigenvalues drop sharply at the beginning and then gradually approach zero after a certain point.

Figure 10a & b. Eigenvalues for TAI sub-indicators



Note: (a) Eigenvalues from traditional Principal Components Analysis (Scree plot); (b) Bootstrapped eigenvalues (1000 samples randomly selected with replacement).

The correlation coefficients between the principal components Z and the variables x are called *component loadings*, $r(Z_j, x_i)$. In case of uncorrelated variables *x*, the loadings are equal to the weights a_{ij} given in equation (1). Analogous to Pearson's r, the squared loading is the percent of variance in that variable explained by the principal component. The *component scores* are the scores of each case (country in our example) on each principal component. The component score for a given case for a principal component is calculated by taking the case's standardised value on each variable, multiplying by the corresponding loading of the variable for the given principal component factor, and summing these products.

Table 7 presents the component loadings for the TAI sub-indicators. High and moderate loadings (>0.50) indicate how the sub-indicators are related to the principal components. It can be seen that with the exception of PATENTS and ROYALTIES, all the other sub-indicators are entirely accounted for by one principal component alone and that the high and moderate loadings are all found in the first four principal components. An undesirable property of these components is that two sub-indicators are related strongly to two principal components.

Table 7. Component loadings for TAI sub-indicators

	1	2	3	4	5	6	7	8
PATENTS	-0.11	**-0.75**	0.13	**0.60**	-0.10	-0.12	-0.17	0.05
ROYALTIES	**-0.56**	-0.48	0.22	**-0.54**	0.27	-0.17	-0.04	0.10
INTERNET	**-0.92**	0.21	0.02	-0.10	0.04	0.11	-0.27	-0.13
EXPORTS	0.35	**-0.85**	0.01	-0.13	0.11	0.35	0.06	-0.08
TELEPHONES	**-0.76**	-0.39	-0.16	-0.16	-0.41	-0.16	0.16	-0.09
ELECTRICITY	**-0.91**	0.13	0.01	0.07	-0.19	0.30	0.04	0.16
SCHOOLING	**-0.74**	0.11	0.37	0.39	0.33	-0.02	0.20	-0.07
ENROLMENT	-0.36	-0.12	**-0.87**	0.15	0.26	-0.03	0.02	0.02

Note: Extraction method: PCA. Loadings greater than 0.5 (absolute values) are highlighted, n=23 countries.

The question of how many principal components should be retained in the analysis without losing too much information and how the interpretation of the components might be improved are addressed in the following section on Factor Analysis.

Factor analysis

Factor analysis (FA) is similar to PCA. It aims to describe a set of Q variables $x_1, x_2, ..., x_Q$ in terms of a smaller number of m factors, and highlight the relationship between these variables. However, whereas the PCA simply is based on linear data combinations, the FA is based on a rather special model (Spearman, 1904). Contrary to the PCA, the FA model assumes that the data is based on the underlying factors of the model, and that the data variance can be decomposed into that accounted for by common and unique factors. The model is given by:

$$
\begin{aligned}
x_1 &= \alpha_{11}F_1 + \alpha_{12}F_2 + ... + \alpha_{1m}F_m + e_1 \\
x_2 &= \alpha_{21}F_1 + \alpha_{22}F_2 + ... + \alpha_{2m}F_m + e_2 \\
&... \\
x_Q &= \alpha_{Q1}F_1 + \alpha_{Q2}F_2 + ... + \alpha_{Qm}F_m + e_Q
\end{aligned}
\qquad \textbf{(4)}
$$

where x_i (i=1,…,Q) represents the original variables but standardized with zero mean and unit variance; $\alpha_{i1}, \alpha_{i2}, ..., \alpha_{im}$ are the factor loadings related to the variable X_i; $F_1, F_2, ..., F_m$ are m uncorrelated common factors, each with zero mean and unit variance; and e_i are the Q specific factors supposed independently and identically distributed with zero mean. There are several approaches to deal with the model given in equation (4), e.g. communalities, maximum likelihood factors, centroid method, principal axis method, etc. The most common is the use of PCA to extract the first m principal components and consider them as factors and neglect the remaining. Principal components factor analysis is most preferred in the development of composite indicators, e.g., Product Market Regulation Index (Nicoletti et al. 2000), as it has the virtue of simplicity and allows the construction of weights representing the information content of sub-indicators. Notice however that different extraction methods supply different values for the factors thus for the weights, influencing the score of the composite and the corresponding country ranking.

On the issue of how factors should be retained in the analysis without losing too much information, methodologists' opinions differ. The decision of when to stop extracting factors basically depends on when there is only very little "random" variability left, and it is rather arbitrary. However, various guidelines ("stopping rules") have been developed, roughly in the order of frequency of their use in social science (Dunteman, 1989: 22-3) (**Box 4**).

Box 4. A sample of "stopping rules"

Kaiser criterion. Drop all factors with eigenvalues below 1.0. The simplest justification to this rule is that it doesn't make sense to add a factor that explains less variance than is contained in one sub-indicator. According to this rule, 3 factors should be retained in the analysis of the TAI example, although the 4th factor follows closely with an eigenvalues of 0.90.

Scree plot. This method proposed by Cattell plots the successive eigenvalues, which drop off sharply and then tend to level off. It suggests retaining all eigenvalues in the sharp descent before the first one on the line where they start to level off. This approach would result in retaining 3 factors in the TAI example (

Figure a).

Variance explained criteria. Some researchers simply use the rule of keeping enough factors to account for 90% (sometimes 80%) of the variation. The first 4 factors account for 87.2% of the total variance.

Joliffe criterion. Drop all factors with eigenvalues under 0.70. This rule may result in twice as many factors as the Kaiser criterion, and it is less often used. In the present case study, this criterion would have lead to the selection of 4 factors.

Comprehensibility. Though not a strictly mathematical criterion, there is much to be said for limiting the number of factors to those whose dimension of meaning is readily comprehensible. Often this is the first two or three.

A relatively recent method for deciding on the number of factors to retain combines the bootstrapped eigenvalues and eigenvectors (Jackson, 1993; Yu *et al*., 1998). Based on a combination of the Kaiser criterion and the bootstrapped eigenvalues, we should consider the first 4 factors in the TAI example. In light of the above analysis, we retain the first four principal components as identified by the bootstrap eigenvalue approach combined with the Kaiser criterion. This choice implies a greater willingness to overstate the significance of the fourth component and be in line with the idea that there are four main categories of technology achievement indicators.

After choosing the number of factors to keep, rotation is a standard step performed to enhance the interpretability of the results (Kline, 1994). The sum of eigenvalues is not affected by rotation, but changing the axes, will alter the eigenvalues of particular factors and will change the factor loadings. There are various rotational strategies that have been proposed. The goal of all of these strategies is to obtain a clear pattern of loadings. However, different rotations imply different loadings, and thus different meanings of principal components - a problem some cite as a drawback to the method. The most common rotation method is the “varimax rotation”.

Table 8 presents the factor loadings for the first factors in the TAI example. Note that the eigenvalues have been affected by the rotation. The variance accounted for by the rotated components is spread more evenly than for the unrotated components. The first four factors account now for 87% of the total variance and are not sorted into descending order according to the amount of the original’s dataset variance explained. The first factor has high positive coefficients (loadings) with Internet (0.79), electricity (0.82) and schooling (0.88). Factor 2 is mainly dominated by patents and exports, whilst enrolment is exclusively loaded on Factor 3. Finally, Factor 4 is formed by royalties and telephones. Yet, despite the rotation of factors, the sub-indicator of exports has sizeable loadings in both Factor 1 (negative loading) and Factor 2 (positive loading). A meaningful interpretation of the factors is not straightforward. Furthermore, the statistical treatment of the eight sub-indicators results in different groups (factors) than the conceptual ones (see Table A.1 in Appendix).

Table 8. Rotated factor loadings for TAI sub-indicators (method 1)

	Factor 1	Factor 2	Factor 3	Factor 4	Communality
PATENTS	0.07	**0.97**	0.06	0.06	0.95
ROYALTIES	0.13	0.07	-0.07	**0.93**	0.89
INTERNET	**0.79**	-0.21	0.21	0.42	0.89
EXPORTS	-0.64	**0.56**	-0.04	0.36	0.86
TELEPHONES	0.37	0.17	0.38	**0.68**	0.77
ELECTRICITY	**0.82**	-0.04	0.25	0.35	0.85
SCHOOLING	**0.88**	0.23	-0.09	0.09	0.85
ENROLMENT	0.08	0.04	**0.96**	0.04	0.93
Explained variance	2.64	1.39	1.19	1.76	
Cumulative (%)	33	50	65	87	

Note: Extraction method: principal components, varimax normalised rotation. Positive loadings greater than 0.5 are highlighted.

Another method of extracting factors that deals with the uncorrelation issue of the specific factors would have given different results. **Table 9** presents the rotated factor loadings of the four factors for the TAI case study (extraction method: principal factors maximum likelihood). For instance, electricity and schooling are not loaded any more both on F1, but electricity is loaded on F4 and schooling on F3. There is 76% variance that is common in the sub-indicators set and expressed by the four rotated common factors. In contrast, the total variance explained in the previous analysis by the four rotated principal components was much higher (87%). The commonalties for seven sub-indicators are greater than 0.66, with the exception of enrolment for which the communality is only 0.15, which indicates that enrolment does not move with the other sub-indicators in the dataset, and therefore it is not well-represented by the four common factors. This conclusion does not depend on the factor analysis method applied, as it has been confirmed by different methods (centroid method, principal axis method).

Table 9. Rotated factor loadings for TAI sub-indicators (method 2)

	Factor 1	Factor 2	Factor 3	Factor 4	Communality
PATENTS	0.01	0.11	**0.88**	0.13	0.80
ROYALTIES	**0.96**	0.14	0.09	0.18	0.99
INTERNET	0.31	**0.56**	-0.29	**0.60**	0.86
EXPORTS	0.29	-0.45	**0.58**	-0.14	0.65
TELEPHONES	0.41	0.13	0.18	**0.73**	0.75
ELECTRICITY	0.13	0.57	-0.13	**0.73**	0.89
SCHOOLING	0.14	**0.95**	0.10	0.14	0.95
ENROLMENT	-0.01	0.03	0.03	0.39	0.15
Explained Variance	1.31	1.80	1.27	1.67	
Cumulative (%)	16	39	55	76	

Note: Extraction method: principal factors maximum likelihood, varimax normalised rotation.

To sum up the steps of PCA/FA as exploratory analysis method:

1. Calculate the covariance/correlation matrix: if the correlation between sub-indicators is small, it is unlikely that they share common factors.

2. Identify the number of factors that are necessary to represent the data and the method for calculating them.

3. Rotate factors to enhance their interpretability (by maximising loading of sub-indicators individual factors).

Cronbach Coefficient Alpha

The Cronbach Coefficient Alpha, c-alpha henceforth, (Cronbach, 1951) is the most common estimate of internal consistency of items in a model or survey – Reliability/Item Analysis (e.g. Boscarino et al., 2004; Raykov, 1998; Cortina, 1993; Feldt et al., 1987; Green et al., 1977; Hattie, 1985; Miller, 1995). It assesses how well a set of items (in our terminology sub-indicators) measures a single unidimensional object (e.g. attitude, phenomenon etc.).

Cronbach's Coefficient Alpha can be defined as:

$$\alpha_c = \left(\frac{Q}{Q-1}\right)\frac{\sum_{i\neq j} cov(x_i, x_j)}{var(x_o)} = \left(\frac{Q}{Q-1}\right)\left(1-\frac{\sum_j var(x_j)}{var(x_o)}\right) \quad c = 1,..,M;\ i,j = 1,..,Q \tag{5}$$

where M indicates the number of countries considered, Q the number of sub-indicators available, and $x_o = \sum_{q=1}^{Q} x_j$ is the sum of all sub-indicators. C-alpha measures the portion of total variability of the sample of sub-indicators due to the correlation of indicators. It increases with the number of sub-indicators and with the covariance of each pair. If no correlation exists and sub-indicators are independent then C-alpha is equal to zero, while if sub-indicators are perfectly correlated the C-alpha is equal to one.

C-alpha is not a statistical test, but a coefficient of reliability based on the correlation between sub-indicators. That is if the correlation is high, then there is evidence that the sub-indicators are measuring the same underlying construct. Therefore a high c-alpha, or equivalently a high “reliability”, indicates that the sub-indicators measure well the latent phenomenon. Although widely interpreted as such, strictly speaking c-alpha is *not a measure of unidimensionality*. A set of sub-indicators can have a high alpha and still be multidimensional. This happens when there are separate clusters of sub-indicators (separate dimensions) which inter-correlate highly, even though the clusters themselves are not highly correlated. An issue is how large the c-alpha must be. Nunnally (1978) suggests 0.7 as an acceptable reliability threshold. Yet, some authors use .75 or .80 as cut-off value, while others are as lenient as .60. In general this varies by discipline.

If the variances of the sub-indicators vary widely, like in our test case, a standard practice is to standardise the sub-indicators to a standard deviation of 1 before computing the coefficient alpha. In our notation this would mean substituting x_i with I_i. The c-alpha is .70 for the dataset of the 23 countries, which is equal to the Nunnally's cutoff value. An interesting exercise is to determine how the c-alpha varies with the deletion of each sub-indicator at a time. This helps to detect the existence of clusters of sub-indicators, thus it is useful to determine the nested structure of the composite. If the reliability coefficient increases after deleting a sub-indicator from the scale, one can assume that the sub-indicator is not correlated highly with other sub-indicators in the scale.

Table 10 presents the values for the Cronbach coefficient alpha and the correlation with the total after deleting one sub-indicator at-a-time. Telephones has the highest variable-total correlation and if deleted the coefficient alpha would be as low as 0.60. If exports were to be deleted from the set then the value of standardised coefficient alpha will increase from the current .70 to .77. Note that the same sub-indicator has the lowest variable-total correlation value (-.108). This indicates that exports is not measuring the same construct as the rest of the sub-indicators are measuring. Note also, that the factor analysis in the previous section had indicated enrolment as the sub-indicator that shares the least amount of common variance with the other sub-indicators. Although both factor analysis and the Cronbach coefficient alpha are based on correlations among sub-indicators, their conceptual framework is different.

Table10. Cronbach coefficient alpha results for TAI sub-indicators

Deleted sub-indicator	Correlation with total	Cronbach coefficient alpha
PATENTS	0.261	0.704
ROYALTIES	0.527	0.645
INTERNET	0.566	0.636
EXPORTS	-0.108	0.774
TELEPHONES	0.701	0.603
ELECTRICITY	0.614	0.624
SCHOOLING	0.451	0.662
ENROLMENT	0.249	0.706

Note: Cronbach coefficient alpha results for the 23 countries after deleting one sub-indicator (standardised values) at a time.

Cluster analysis

Cluster analysis (CLA) is a collection of algorithms to classify objects such as countries, species, and individuals. (Anderberg 1973, Massart and Kaufman 1983). The classification aims to reduce the dimensionality of a dataset by exploiting the similarities/dissimilarities between cases. CLA techniques can be hierarchical, if the classification has an increasing number of nested classes, e.g., *tree clustering*; or non-hierarchical when the number of clusters is decided ex ante e.g., the *k-means clustering*. However, care should be given that classes are meaningful, and are not arbitrary or artificial.

Homogeneous and distinct groups could be delineated based upon assessment of distances or as in the case of Ward's method, an F-test (Davis, 1986). A distance measure is an appraisal of the degree of similarity or dissimilarity between cases in the set. A small distance is equivalent to a large similarity. It can be based on a single dimension or on multiple dimensions, for example countries in TAI example can be evaluated according to the TAI composite indicator or they can be evaluated according to all single sub-indicators. Some of the most common distance measures are listed in **Table 11** including Euclidean and non-Euclidean distances[4].

Table 11. Distance measures for TAI sub-indicators

Euclidean	$D(x,y)=\left(\frac{\sum_{i=1}^{N_d}(x_i-y_i)^2}{N_d}\right)^{1/2}$	This is the geometric distance in a multidimensional space and is usually computed from raw data (prior to any normalisation). This measure is not affected by the addition of new objects such as outliers. However, it is highly affected by the difference in scale, e.g., whether the same object is measured in centimetres or in meters the *D(x,y)*
Squared Euclidean	$D(x,y)=\frac{\sum_{i=1}^{N_d}(x_i-y_i)^2}{N_d}$	This measure places progressively greater weight on objects that are further apart. Usually this is computed from raw data and shares the same advantages and disadvantages of the Euclidean distance.
City-block[5] (Manhattan)	$D(x,y)=\frac{\sum_{i=1}^{N_d}\lvert x_i-y_i\rvert}{N_d}$	This distance is the average of distances across dimensions and it yields similar results to the Euclidean distance. In this measure the effect of outliers is less pronounced, since it is not squared.
Chebychev	$D(x,y)=Max\lvert x_i-y_i\rvert$	This measure is mostly used when one wants to define objects as "different" if they are different in any one of the dimensions.
Power	$D(x,y)=\left(\frac{\sum_{i=1}^{N_d}(x_i-y_i)^p}{N_d}\right)^{1/r}$	This distance measure is useful when one wants to increase or decrease the progressive weight placed on one dimension, for which the respective objects are very different. The parameters *r* and *p* are user-defined, such that *p* controls the progressive weights placed on differences on individual dimensions, and *r* controls the progressive weight placed on larger differences between objects. Note that for *p* = *r* = 2 corresponds to the Euclidean distance.
Percent disagreement	$D(x,y)=\frac{number\ of\ x_i\neq y_i}{N_d}$	This measure is useful if the data are categorical in nature.

The next step is to choose the clustering algorithm, i.e. the rules, which govern how distances are measured between clusters. There are many methods available. The selection criteria could differ and hence different classifications may be obtained for the same data, even using the same distance measure. The most common linkage rules are (Spath, 1980):

- **Single linkage** (nearest neighbour). The distance between two clusters is determined by the distance between the two closest elements in the different clusters. This rule called also single linkage, produces clusters chained together by single objects.

- **Complete linkage** (farthest neighbour). The distance between two clusters is determined by the greatest distance between any two objects belonging to different clusters. This method usually performs well when objects naturally form distinct groups.

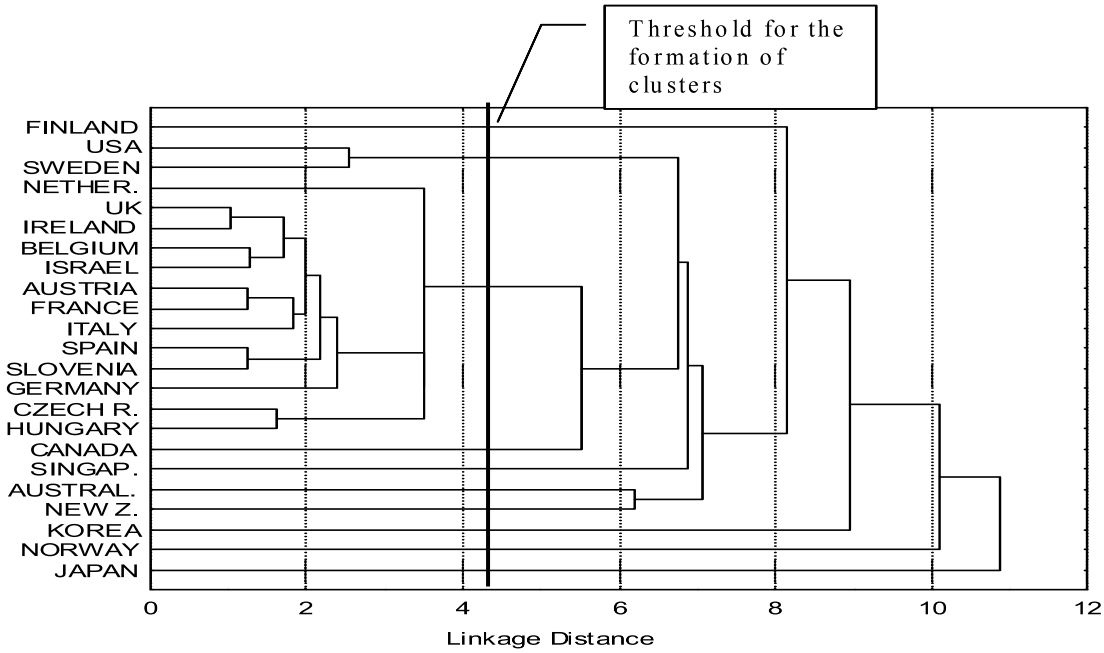
- **Unweighted pair-group average**. The distance between two clusters is calculated as the average distance between all pairs of objects in the two clusters. This method usually performs well when objects naturally form distinct groups. A variation of this method is using the centroid of a cluster--the distance is the average point in the multidimensional space defined by the dimensions.

- **Weighted pair-group average**. Similar to the unweighted pair-group average (centroid included) except that the size of the cluster, i.e. the number of objects contained, is used as weight for the average distance. This method is useful when cluster sizes are very different.

- **Ward's method** (Ward, 1963). Cluster membership is determined by calculating the variance of elements, i.e., the sum of the squared deviations from the mean of the cluster. An element will belong to the cluster if it produces the smallest possible increase in the variance.

Figure 11 shows the country clusters based on the technology achievement sub-indicators using tree clustering (hierarchical) with single linkage and squared Euclidean distances. Similarity between countries belonging to the same cluster decreases as the linkage distance increases. One of the biggest problems with CLA is identifying the optimum number of clusters. As the amalgamation process continues increasingly dissimilar clusters must be fused, i.e. the classification becomes increasingly artificial. Deciding upon the optimum number of clusters is largely subjective, although looking at the plot of linkage distance across fusion steps may help (Milligan and Cooper, 1985).

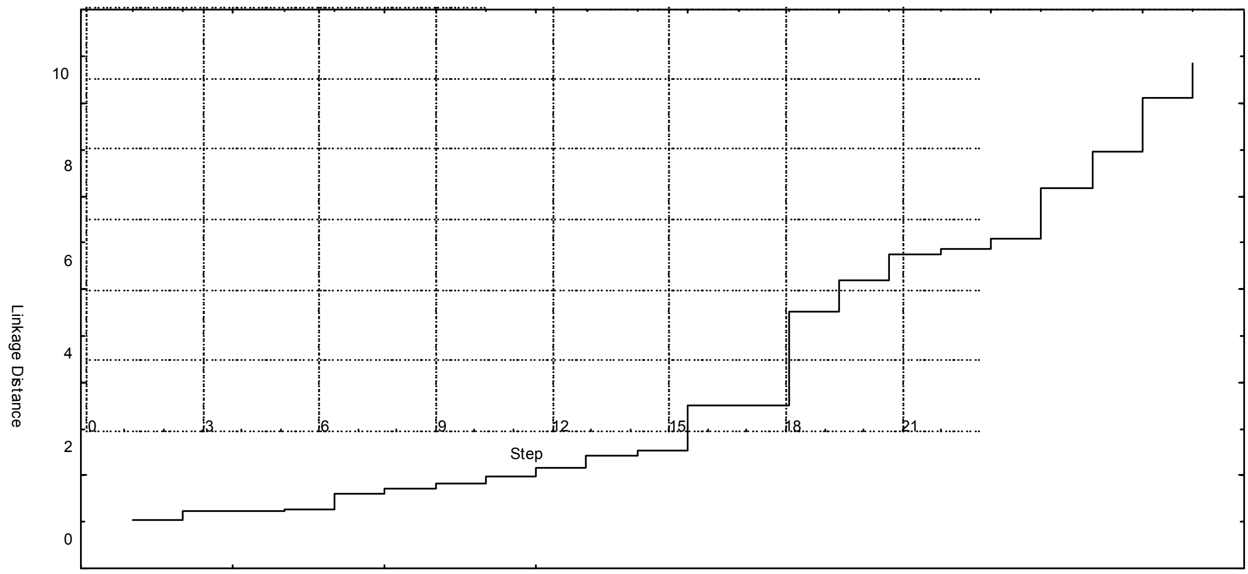
Figure 11. Country clusters for TAI sub-indicators



Note: Standardised data. Type: Hierarchical, single linkage, squared Euclidean distances.

Sudden jumps in the level of similarity (abscissa) could indicate that dissimilar groups or outliers are fused. Such a plot is presented in **Figure 12**, where the greatest dissimilarity among the 23 countries in the TAI example is found at a linkage distance close to 4.0, which indicates that the data are best represented by ten clusters: Finland; Sweden and USA; the group of countries located between the Netherlands and Hungary; Canada; Singapore; Australia; New Zealand; Korea; Norway; and Japan. Note that the most dissimilar are Korea, Norway and Japan, which are aggregated only at the very end of the analysis. Besides, this result does not fully correspond to the division in laggard, average and leading countries resulting from the standard aggregation methods. Japan, in fact, would be in the group of leading countries, together with Finland, Sweden, USA, while Hungary, Czech Republic, Slovenia and Italy would be the laggards, far away from the Netherlands, USA or Sweden (see Table 30).

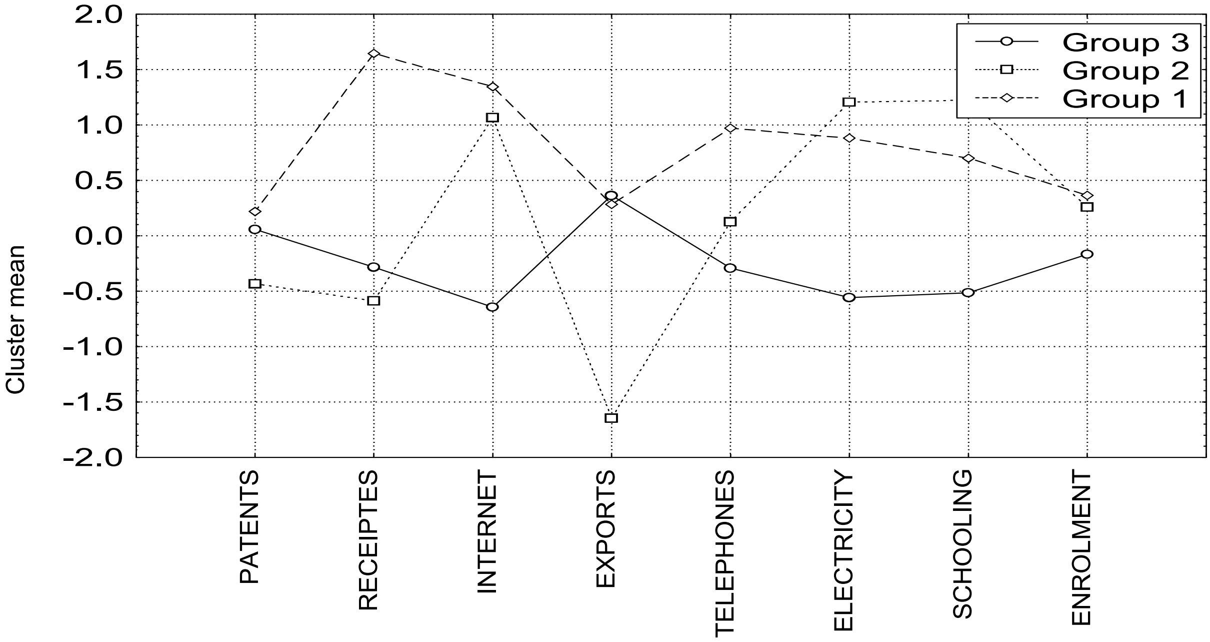
Figure 12. Linkage distance vs fusion step in TAI hierarchical cluster



A non-hierarchical method of clustering, different from the *joining* or *tree* clustering shown below, is the k-means clustering (Hartigan, 1975). This method is useful when the aim is to divide the sample in *k* clusters of greatest possible distinction. The parameter *k* is decided by the analyst, for example we may decide to cluster the 23 countries in the TAI example into 3 groups, e.g., leaders, potential leaders, and dynamic adopters. The k-means algorithm will supply 3 clusters, as distinct as possible, by analysing the variance of each cluster. This algorithm can be applied with continuous variables, yet it can be also modified to accommodate for other types of variables. The algorithm starts with *k* random clusters and moves the objects in and out the clusters with the aim of (i) minimising the variance of elements within the clusters, and (ii) maximise the variance of the elements outside the clusters.

A line graph of the means across clusters is displayed in **Figure 13**. This plot could be very useful in summarising the differences in the means between clusters. It is shown for example that the main difference between the *leaders* and the *potential leaders* (**Table 12**) is on receipts and exports. At the same time, the *dynamic adopters* are lagging behind the *potential leaders* due to their lower performance on Internet, electricity and schooling. They are, however, performing better on exports. Two of the sub-indicators, patents and enrolment, are not useful in distinguishing between these 3 groups, as the cluster means are very close.

Figure13. Means plot for TAI clusters



Note: Type: k-means clustering (standardised data).

Table 12. K-means clustering for TAI countries

Group1 (leaders)	**Group 2** (potential leaders)	**Group 3** (dynamic adopters)
Finland	Canada	Japan
USA	Australia	Korea
Sweden	Norway	UK
Netherlands	New Zealand	Singapore
		Germany
		Ireland
		Belgium
		Austria
		France
		Israel
		Spain
		Italy
		Czech Rep.
		Hungary
		Slovenia

Finally, expectation maximisation (EM) clustering extends the simple k-means clustering in two ways. First, instead of clustering the objects by maximising the differences in means for continuous variables, EM clusters membership on the basis of probability distributions--each observation will belong to each cluster with a certain probability. EM estimates mean and standard deviation of each cluster so as to maximises the overall likelihood of the data, given the final clusters (Binder, 1981). Second, unlike k-means, EM can be applied to both continuous and categorical data.

Ordinary significance tests are not valid for testing differences between clusters, as clusters are formed to be as much separated as possible. Thus the assumption of usual tests--parametric or non-parametric--is violated (see Hartigan 1975). As final remark a warning: CLA will always produce a grouping, this means that clusters may or may not prove useful for classifying objects depending upon the objectives of the analysis. For example, if grouping zip code areas into categories based on age, gender, education and income discriminates between wine drinking behaviours, then this would be useful information only if the aim of the CLA was that of establishing a wine store in new areas. Furthermore, CLA methods are not clearly established, there are many options, all giving very different results (Everitt, 1979).

Various alternative methods combining cluster analysis and the search for a low-dimensional representation have been proposed, and focus on multidimensional scaling or unfolding analysis (e.g., Heiser, 1993, De Soete and Heiser, 1993). A method that combines k-means cluster analysis with aspects of Factor Analysis and PCA is presented by Vichi and Kiers (2001). A discrete clustering model together with a continuous factorial one are fitted simultaneously to two-way data, with the aim to identify the best partition of the objects, described by the best orthogonal linear combinations of the variables (factors) according to the least-squares criterion.

This methodology named factorial k-means analysis has a wide range of applications since it reaches a double objective: data reduction and synthesis simultaneously in direction of objects and variables. Originally applied to short-term macroeconomic data, factorial k-means analysis has a fast alternating least-squares algorithm that extends its application to large data sets, *i.e.*, multivariate data sets with >2 variables. The methodology can therefore be recommended as an alternative to the widely used tandem analysis that sequentially performs PCA and CLA..

IMPUTATION OF MISSING DATA

The literature on the analysis of missing data is extensive and in rapid development. This section covers the main methods. More comprehensive surveys can be found in Little and Rubin (2002), Little (1997) and Little and Schenker (1994).

Single imputation

Imputations are means or draws from a predictive distribution of the missing values (Little and Rubin, 2002). The predictive distribution must be generated by employing the observed data either through implicit or explicit modelling:

Implicit modelling. The focus is on an algorithm, with implicit underlying assumptions that need to be verified whether they are reasonable and fit to the issue under consideration. The danger of this type of modelling missing data is to consider the resulting data set as complete, and forget that an imputation has been done. Implicit modelling includes:

- **Hot deck imputation.** Fill in blanks cells with individual data, drawn from "similar" responding units, e.g. missing values for individual income may be replaced with the income of another respondent with similar characteristics, e.g., age, sex, race, place of residence, family relationships, job, etc.

- **Substitution.** Replace non-responding units with units not selected into the sample, e.g. if a household cannot be contacted, then a previously non-selected household in the same housing block is selected.

- **Cold deck imputation.** Replace the missing value with a value from an external source, e.g. from a previous realisation of the same survey.

Explicit modelling. The predictive distribution is based on a formal statistical model where the assumptions are made explicitly, such as:

- **Unconditional mean/median/mode imputation.** The sample mean (median, mode) of the recorded values for the given sub-indicator replaces the missing values.

- **Regression imputation**. Missing values are substituted by the predicted values obtained from regression. The dependent variable of the regression is the sub-indicator hosting the missing value, and the regressor(s) is (are) the sub-indicator(s), showing a strong relationship with the dependent variable, i.e., usually a high degree of correlation.

- **Expectation Maximisation (EM) imputation.** This model focuses on the interdependence between model parameters and the missing values. The missing values are substituted by estimates obtained through an iterative process. First, one predicts the missing values based on initial estimates of the model parameter values. These predictions are then used to update the parameter values, and the process is repeated. The sequence of parameters converges to maximum-likelihood estimates, and the time to convergence depends on the proportion of missing data and the flatness of the likelihood function.

If simplicity is its main appeal, an important limitation of the single imputation method is its systematic underestimation of the variance of the estimates (with some exceptions for the EM method where the bias depends on the algorithm used to estimate the variance). Therefore, this method does not fully assess the implications of imputation or the robustness of the composite index derived from the imputed dataset.

Unconditional mean imputation

Let X_q be the random variable associated to the sub-indicator q, with $q=1,...,Q$, and $x_{q,c}$ the observed value of X_q for country c, with $c=1,..,M$. Let m_q be the number of recorded or non-missing values on X_q, and $M-m_q$ the number of missing values. The unconditional mean is then given by:

$$\bar{x}_q = \frac{1}{m_q} \sum_{recorded} x_{q,c} \qquad (6)$$

Similarly, the median[6] and the mode[7] of the distribution could be calculated on the available sample and to substitute missing values.[8] By "filling in" blank spaces with the sample mean, the imputed value becomes a biased estimator of the population mean, even in the case of MCAR mechanisms, and the sample variance underestimates true variance, underestimating the uncertainty on the composite due to the imputation.

Regression imputation

Suppose a set of $h-1$<Q fully observed sub-indicators $(x_1,...,x_{h-1})$ and a sub-indicator x_h only observed for r countries, but missing for the remaining $M-r$ countries. Regression imputation computes the regression of x_h on $(x_1,...,x_{h-1})$ using r complete observations, and impute the missing values as prediction from the regression[9]:

$$\hat{x}_{ih} = \hat{\beta}_0 + \sum_{j=1}^{h-1} \hat{\beta}_j x_{ij} \quad i = 1,..,M-r \qquad (7)$$

In general, the strategy to define the 'best' regression is a two step procedure. First, all different subsets of predictors are adopted in a multiple regression manner. Then, the best subset(s) is determined using the following criteria:[10]

- the value of R2
- the value of the residual mean square RMS
- the value of Mallows' Ck
- stepwise regression

A variation of the regression approach is the stochastic regression approach that imputes a conditional draw instead of imputing the conditional mean:

$$\hat{x}_{ih} = \hat{\beta}_0 + \sum_{j=1}^{h-1} \hat{\beta}_j x_{ij} + \varepsilon_i \quad i = 1,..,M-r \qquad (7^*)$$

where, ε_i is a random variable $N(0,\hat{\sigma}^2)$ and $\hat{\sigma}^2$ is the residual variance from the regression of x_h on $(x_1,...,x_{h-1})$ based on the r complete cases.

A key problem for both approaches is again the underestimation of the standard errors, although stochastic regression ameliorates the distortions. Hence, the inference based on the entire dataset, including the imputed data, does not fully count for imputation uncertainty. The result is that *p-values* of tests are too small and confidence intervals too narrow. Replication methods and multiple imputation could correct the loss of precision of simple imputation.

What if the variable with missing observations is categorical? Regression imputation is still possible but adjustment using, *e.g*., rounding of the predictions or a logistic, ordinal or multinomial logistic regression models, is required. For nominal variables, frequency statistics such as the mode or hot- and cold-deck imputation methods might be more appropriate.

Expected maximisation imputation

Suppose that X denotes the matrix of data. In the likelihood based estimation, the data is assumed to be generated by a model, described by a probability or density function $f(X \mid \theta)$, where θ is the unknown parameter vector lying in the parameter space Ω_θ (e.g. the real line for means, the positive real line for variances and the interval [0,1] for probabilities). The probability function captures the relationship between the data set and the parameter of the model, and describes the probability of observing a dataset for a given $\theta \in \Omega_\theta$. Since θ is unknown, while the data set is known, it makes sense to reverse the argument, and look for the probability, or the likelihood, of observing a certain θ given X. Therefore, given X, the likelihood function $L(\theta \mid X)$ is any function of $\theta \in \Omega_\theta$ proportional to $f(X \mid \theta)$:

$$L(\theta \mid X) = k(X) f(X \mid \theta) \tag{8}$$

where, $k(X) > 0$ is a function of X and not of θ. The log-likelihood is then the natural logarithm of the likelihood function. For M independent and identically distributed observations $X = (x_1,...,x_M)^T$ from a normal population with mean μ and variance σ^2 the joint density is

$$f(X \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-M/2} \exp\left(-\frac{1}{2} \sum_{c=1}^{M} \frac{(x_c - \mu)^2}{\sigma^2} \right) \tag{9}$$

For a given sample X the log-likelihood is (ignoring additive constants of function $f(\cdot)$) a function of (μ, σ^2):

$$\begin{aligned} l(\mu, \sigma^2 \mid X) &= \ln[L(\mu, \sigma^2 \mid X)] = \ln[k(X) f(X \mid \mu, \sigma^2)] \\ &= \ln k(X) - \frac{M}{2} \ln \sigma^2 - \frac{1}{2} \sum_{c=1}^{M} \frac{(x_c - \mu)^2}{\sigma^2} \end{aligned} \tag{10}$$

Maximising the likelihood function corresponds to the question of which value of $\theta \in \Omega_\theta$ is mostly supported by a given sampling realisation X. This implies solving the likelihood equation:

$$D_l(\theta \mid X_{obs}) \equiv \frac{\partial \ln L(\theta \mid X_{obs})}{\partial \theta} = 0 \tag{11}$$

When a closed-form solution of equation (11) cannot be found, iterative methods can be applied. The EM algorithm is one of these iterative methods. [11] The issue is that X contains both observable and missing values, i.e. $X = (X_{obs}, X_{mis})$. Thus one has to find both the unknown parameters and the unknown observations of the model.

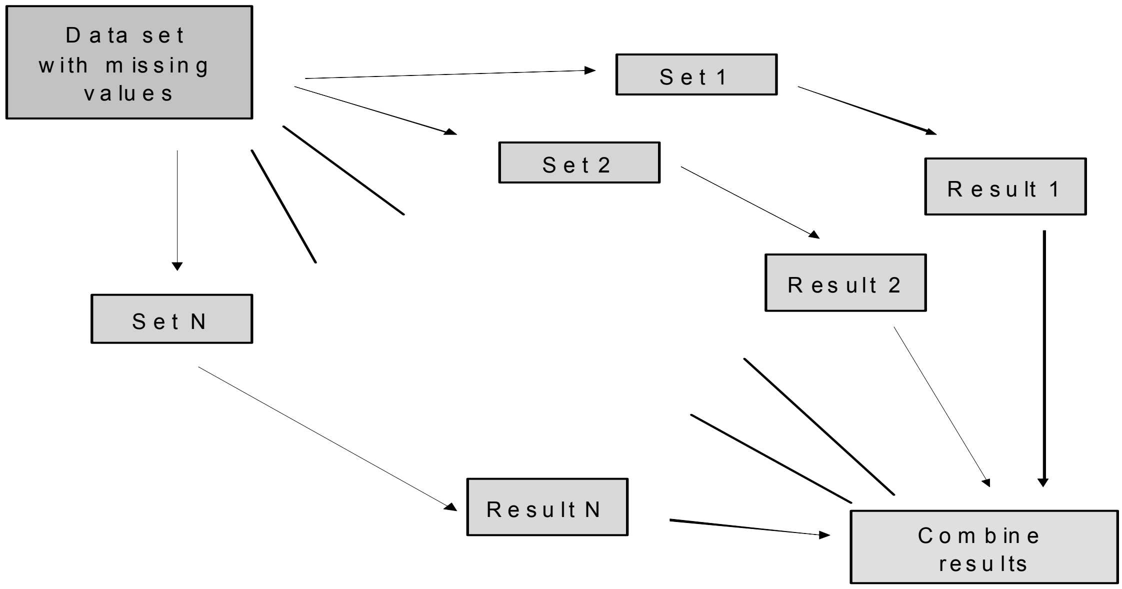
Assuming that missing data are MAR or MCAR[12], the EM consists of two components, the expectation (E) and maximisation (M) steps. Each step is completed once within each algorithm cycle. Cycles are repeated until a suitable convergence criterion is satisfied. The procedure is as follows. First (M) the parameter vector θ is estimated by applying maximum likelihood as if there was no missing data, and Second (E) the expected values of the missing variables are then calculated given the estimate of θ just obtained in the M step. This procedure is repeated until convergence (absence of changes in estimates and in the variance-covariance matrix). Effectively, this process maximises the expectation of the complete data log-likelihood in each cycle conditional on the observed data and parameter vector. To start the process, however, one needs an initial estimate of the missing data. This is done by running the first M step on the non-missing observations only, and then predict the missing variables by using the estimate on θ.

The advantage of the EM is its broadness. It can be used for a broad range of problems, e.g. variance component estimation or factor analysis. EM algorithm is also often easy to construct conceptually and practically. Besides, each step has a statistical interpretation and convergence is reliable. The main drawback however, is that convergence may be very slow, when a large fraction of information is missing (if there was no missing information, convergence would be immediate). The user should also be careful that the maximum found is indeed a global maximum and not a local one. To test this, different initial starting values for theta can be used.

Multiple imputation

Multiple imputation (MI) is a general approach that does not require a specification of parameterised likelihood for all data (**Figure 14)**. The imputation of missing data is performed with a random process that reflects uncertainty. Imputation is done N times, to create N "complete" data sets. On each data set the parameters of interest are estimated, together with their standard errors. Average (mean or median) estimates are combined using the N sets and between and within imputation variance is calculated.

Figure 14. Logic of multiple imputation

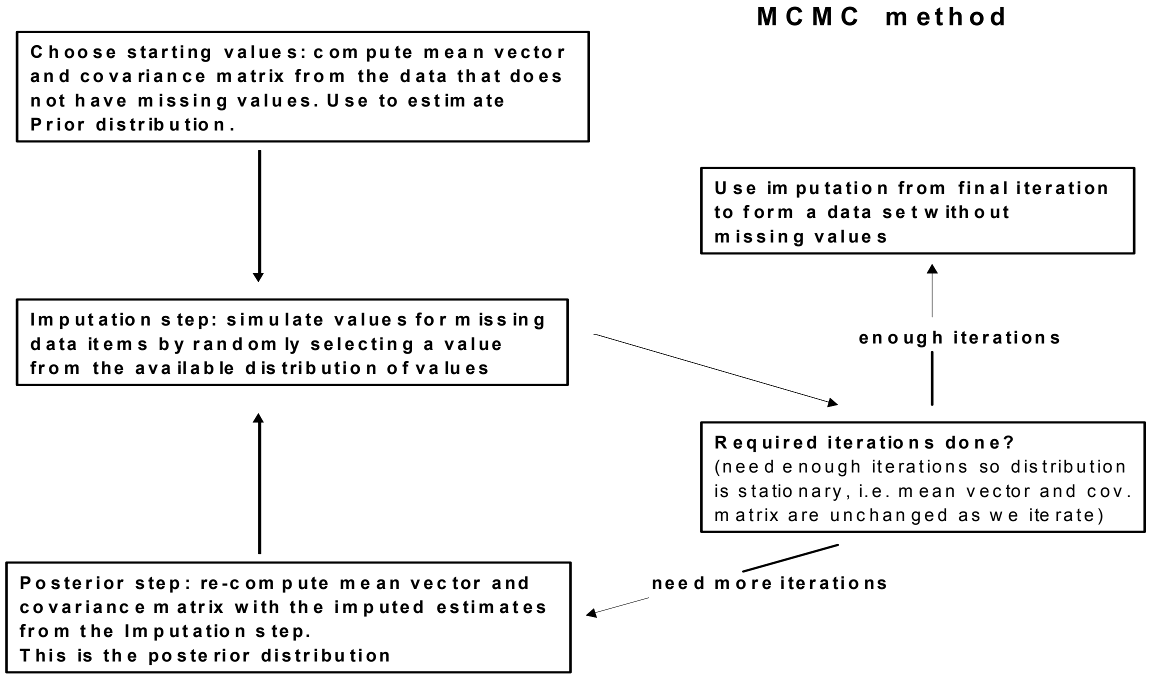


Any "proper" imputation method can be used in multiple imputation. For example, one could use regression imputation repeatedly, drawing *N* values of the regression parameters using the variance matrix of estimated coefficients. However, one of the most general models is the *Markov Chain Monte Carlo* (MCMC) method. MCMC is a sequence of random variables, in which the distribution of the actual element depends on the value of the previous one. It assumes that data are drawn from a multivariate Normal distribution and requires MAR or MCAR assumptions.

The theory of MCMC is most easily understood using Bayesian methodology (**Figure 15**). Denote the observed data as X_{obs}, and the complete dataset as X=(X_{obs}, X_{mis}), where X_{mis} is to be filled in via multiple imputation. If the distribution of X_{mis}, with parameter vector θ, were known then we could impute X_{mis} by drawing from the conditional distribution $f(X_{mis}|X_{obs}, \theta)$. However, since θ is unknown, we shall estimate it from the data, yielding $\hat{\theta}$, and use the distribution $f(X_{mis}|X_{obs} \hat{\theta})$. Since $\hat{\theta}$ is itself a random variable, we must also take its variability into account in drawing imputations.

The missing data generating process may also depend on additional parameters φ but if φ and θ are independent, the process is called ignorable and the analyst can concentrate on modelling the missing data given the observed data and θ**.** If the two processes are not independent, then we have a non-ignorable missing data generating process, which cannot be solved adequately without making assumptions on the functional form of the interdependency.

Figure 15. Markov Chain Monte Carlo iimputation method



Source: Rearranged from K. Chantala and C. Suchindran,
(http://www.cpc.unc.edu/services/computer/presentations/mi_presentation2.pdf)

In Bayesian terms, θ is a random variable, whose distribution depends on the data. The first step for its estimation is to obtain the posterior distribution of θ from the data. Usually this posterior is approximated by a normal distribution. After formulating the posterior distribution of θ, the following imputation algorithm can be used:

- Draw θ^* from the posterior distribution of θ, f(θ|Y, Xobs) where Y denotes exogenous variables that may influence θ.
- Draw Xmis from f(Xmis|Y, Xobs, θ^*)
- Use the completed data X and the model to estimate the parameter of interest (e.g. the mean) β^* and its variance V(β^*) (within-imputation variance).

These steps are repeated independently N times, resulting in β_n^*, $V(\beta_n^*)$, $n=1,\ldots,N$. Finally, the N imputations are combined. A possible *combination* is the mean of all individual estimates (but also the median can be used):

$$\beta^* = \frac{1}{N}\sum_{n=1}^{N}\beta_n^* \qquad (12)$$

This *combination* will be the value that fills in the blank space in the dataset. The total variance is obtained as a weighted sum of the *within-imputation* variance and the *between-imputations* variance:

$$V^* = \overline{V} + \frac{N+1}{N}B \qquad (13)$$

where the mean of the *within-imputation* variances is

$$\overline{V} = \frac{1}{N}\sum_{n=1}^{N} V\left(\beta_n^*\right) \tag{14}$$

and the *between-imputations* variance is given by

$$B = \frac{1}{N-1}\sum_{n=1}^{N}\left(\beta_n^* - \beta^*\right)\left(\beta_n^* - \beta^*\right)' \tag{15}$$

Confidence intervals are obtained by taking the overall estimate plus or minus a multiple of standard error, where that number is a quantile of Student's t-distribution with degrees of freedom:

$$df = (N-1)\left(1+\frac{1}{r}\right)^2 \tag{16}$$

where *r* is the *between-to-within* ratio.

$$r = \left(1+\frac{1}{N}\right)\frac{B}{\overline{V}} \tag{17}$$

Based on these variances, one can calculate approximate 95% confidence intervals.

Multiple Imputation method imputes several values (N) for each missing value (from the predictive distribution of the missing data), to represent the uncertainty about which values to impute. The N versions of completed data sets are analysed by standard complete data methods and the results are combined using simple rules to yield single combined estimates (e.g., MSE, regression coefficients), standard errors, p-values, that formally incorporate missing data uncertainty. The pooling of the results of the analyses performed on the multiple imputed data sets, implies that the resulting point estimates are averaged over the N completed sample points, and the resulting standard errors and p-values are adjusted according to the variance of the corresponding N completed sample point estimates. Thus, the '*between imputation variance*', provides a measure of the extra inferential uncertainty due to missing data, which is not reflected in single imputation).

NORMALISATION

The objective is to identify the most suitable normalization procedures to apply to the problem at hand, taking into account their properties with respect to the measurement units in which the indicators are expressed, and their robustness to possible outliers in the data (Ebert and Welsch, 2004). Different normalization methods will supply different results for the composite indicator. Therefore, overall robustness tests should be carried out to assess their impact on the outcomes

Scale transformation prior to normalisation

Certain normalisation procedures provide the same normalised value of the indicator irrespective of the measurement unit. Applying a normalisation procedure, which is not invariant to changes in the measurement unit, however could result in different outcomes. Below is a simple example with two indicators--temperature and humidity--for two hypothetical countries A and B, in 2003 and 2004. The raw data and normalised composites are given in **Table 13** where the temperature is first expressed in Celsius and then in Fahrenheit. Each indicator is divided by the value of the leading country and aggregated with equal weights. Using Celsius data normalised based on "distance to the best performer", the performance of Country A has increased over time. While the same normalisation and aggregation methods are used, the results in Fahrenheit show a completely different pattern. The composite indicator for country A now decreases over time.

Table 13. Normalisation based on interval scales

Temperature data in Celsius		
	2003	2004
Country A –Temperature (ºC)	35	35.9
Country A –Humidity (%)	75	70
Country B –Temperature (ºC)	39	40
Country B –Humidity (%)	50	45
Normalised data in Celsius		
Country A	0.94872	0.94875
Country B	0.83333	0.82143
Temperature data in Fahrenheit		
Country A –Temperature (F)	95	96.62
Country A –Humidity (%)	75	70
Country B –Temperature (F)	102.2	104
Country B –Humidity (%)	50	45
Normalised data in Fahrenheit		
Country A	0.964775	0.964519
Country B	0.83333	0.82143

The example illustrated so far is a case of *interval scale*, based on a transformation f defined as:

$$f: \quad x \rightarrow y = \alpha x + \beta; \alpha > 0, \beta \neq 0$$

where, the variable x is the temperature expressed in Celsius and y is the temperature expressed in Fahrenheit. Their relationship is given by:

$$y(\mathrm{F}) = \frac{9}{5}x(^{\circ}\mathrm{C}) + 32$$

Another common change of measurement unit is the so-called *ratio scale*, which is based on the transformation:

$$f: \quad x \to y = \alpha x; \alpha > 0 .$$

To give an example, a "length" might be expressed in centimetres (cm) or yards (yd). Their relationship is indeed: 1 yd = 91.44 cm. The normalisation by country leader, not invariant on the 'interval scale', is invariant on the 'ratio scale'. In general, all normalisation methods that are invariant on the 'interval scale', are also invariant on the 'ratio scale'.

Another transformation, which is often used to reduce the skewness of (positive) data, is the logarithmic transformation:

$$f: \quad x \to y = \log(x); x > 0 .$$

When the range of values for the indicator is wide, or it is positively skewed, the log transformation shrinks the right-hand side of the distribution. As values approach zero they are also penalised, given after transformation, they become largely negative. Expressing the weighted variables in a linear aggregation in logarithms is equivalent to the geometric aggregation of the variables without logarithms. The ratio between two weights indicates the percentage improvement in one indicator that would compensate for one percentage point decline in another indicator. This transformation leads to attributing higher weight for a one-unit improvement, starting from a low level of performance, compared to an identical improvement starting from a high level of performance.

The normalisation methods described below are all non-invariant to this type of scale transformation. The user may decide whether or not to use the log transformation before the normalisation, bearing in mind that the normalised data will be affected by the log transformation.

In some circumstances, outliers[13] can reflect the presence of unwanted information. An example is offered in the *Environmental Sustainability Index*, where the variable distributions outside the 2.5 and 97.5 percentile scores are trimmed to partially correct for outliers, as well as to avoid having extreme values overly dominate the aggregation algorithm. That is, any observed value greater than the 97.5 percentile is lowered to match the 97.5 percentile. Any observed value lower than the 2.5 percentile on is raised to the 2.5 percentile. It is advisable to first try to remove outliers, and consequently perform the normalisation, as this latter procedure can be more or less sensitive to outliers.

Standardisation (or z-scores)

For each sub-indicator x_{qc}^{t}, the average across countries $x_{qc=\bar{c}}^{t}$ and the standard deviation across countries $\sigma_{qc=\bar{c}}^{t}$ are calculated. The normalisation formula is: $I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc=\bar{c}}^{t}}{\sigma_{qc=\bar{c}}^{t}}$, so that all the I_{qc}^{t} have similar dispersion across countries. The actual minima and maxima of the I_{qc}^{t} across countries depend on the sub-indicator. For time-dependent studies, in order to assess country performance across years, the average across countries $x^{t_0}_{qc=\bar{c}}$ and the standard deviation across countries $\sigma^{t_0}_{qc=\bar{c}}$ are calculated for a reference year, usually the initial time point, t_0 .

Re-scaling

Each indicator x^t_{qc} for a generic country c and time t is transformed in $I^t_{qc} = \dfrac{x^t_{qc} - min_c(x^t_q)}{max_c(x^t_q) - min_c(x^t_q)}$ where $min_c(x^t_q)$ and $max_c(x^t_q)$ are the minimum and the maximum value of x^t_{qc} across all the countries c at time t. In this way, the normalised indicators I_{qc} have values laying between 0 (laggard, $x^t_{qc} = min_c(x^t_q)$), and 1 (leader, $x^t_{qc} = max_c(x^t_q)$.

The expression $I^t_{qc} = \dfrac{x^t_{qc} - min_c(x^{t_0}_q)}{max_c(x^{t_0}_q) - min_c(x^{t_0}_q)}$ is sometimes used for time-dependent studies. However, if $x^t_{qc} > max_c(x^{t_0}_q)$, the normalised indicator y^t_{qc} would be larger than 1.

Another variant of the re-scaling method is $I^t_{qc} = \dfrac{x^t_{qc} - min_{t\in T}\, min_c(x^t_q)}{max_{t\in T}\, max_c(x^t_q) - min_{t\in T}\, min_c(x^t_q)}$ where the minimum and maximum for each indicator is calculated across countries and time to take into account the evolution of indicators. The normalised indicators, I^t_{qc}, have values between 0 and 1. However, this transformation is not stable, when data for a new time point becomes available. This implies an adjustment of the analysis period T, which may, in turn, affect the minimum and the maximum for some sub-indicators and, hence the values of I^t_{qc} . To maintain comparability between the existing and the new data, the composite indicator for the existing data needs be recalculated.

Distance to a reference

This method takes the ratios of the indicator x^t_{qc} for a generic country c and time t with respect to the sub-indicator $x^{t_0}_{qc=\bar{c}}$ for the reference country at the initial time t_0 .

$$I^t_{qc} = \frac{x^t_{qc}}{x^{t_0}_{qc=\bar{c}}}$$

Using the denominator $x^{t_0}_{qc=\bar{c}}$, the transformation takes into account the evolution of indicators across time; alternatively one can use the denominator $x^t_{qc=\bar{c}}$, with running time t.

A different approach is to consider the country itself as the reference country and calculate the distance in terms of the initial time point as

$$I^t_{qc} = \frac{x^t_{qc}}{x^{t_0}_{qc}} .$$

This approach is used in *Concern about environmental problems* (Parker, 1993) for measuring the concern of the public on certain environmental problems in three countries (Italy, France and the UK) and in the European Union. An alternative distance for the normalisation can be:

$$y_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc=\bar{c}}^{t_0}}{x_{qc=\bar{c}}^{t_0}}$$

which is essentially same as above. Instead of being centred on one, it is centred on zero. In the same way, the reference country can either be the average country, the group leader, or an external benchmark.

Indicators above or below the mean

This transformation considers the indicators that are above and below an arbitrarily defined threshold, *p,* around the mean:

$$I_{qc}^{t} = \begin{cases} 1 & \textbf{\textit{if}} \quad w > (1+p) \\ 0 \quad \textbf{\textit{if}} & (1-p) \leq w \leq (1+p) \\ -1 & \textbf{\textit{if}} \quad w < (1+p) \end{cases}, \text{ where } w = x_{qc}^{t} \big/ x_{qc=\bar{c}}^{t_0}$$

The threshold builds a neutral region around the mean, where the transformed indicator is zero. This reduces the sharp discontinuity from -1 to +1 that exists across the mean value, to two minor discontinuities from -1 to 0 and from 0 to +1 across the thresholds. A larger number of thresholds could be created at different distances from the mean value, which might overlap with the categorical scales. For time-dependent studies to assess country performance over time, the average across countries $x_{qc=\bar{c}}^{t_0}$ is calculated for a reference year (usually the initial time point t_0). An indicator that moves from significantly below the mean to significantly above the threshold in the consecutive year will have a positive effect on the composite.

Methods for cyclical indicators

When indicators are in the form of time series the transformation can be made by subtracting the mean over time $E_t\left(x_{qc}^{t}\right)$ and then by dividing by the mean of the absolute values of the difference from the mean. The normalised series are then converted into index form by adding 100.

$$I_{qc}^{t} = \frac{x_{qc}^{t} - E_t\left(x_{qc}^{t}\right)}{E_t\left(\left|x_{qc}^{t} - E_t\left(x_{qc}^{t}\right)\right|\right)}$$

Percentage of annual differences over consecutive years

Each indicator is transformed using the formula:

$$I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc}^{t-1}}{x_{qc}^{t}} * 100$$

The transformed indicator is dimension-less.

Examples of the above normalisation methods are shown in **Table 14** using the TAI data. The data are sensitive to the choice of the transformation and this might cause problems in terms of loss of the interval level of the information, sensitivity to outliers, arbitrary choice of categorical scores and sensitivity to weighting.

Table 14. Examples of normalisation techniques using TAI data

Country	Mean years of school age 15 and above	Rank high value = top in the list	z-score	re-scaling	distance to reference country ratio c=mean	distance to reference country ratio c=best	distance to reference country ratio c= worst	distance to reference country difference c=mean	distance to reference country difference c=worst	Above/below the mean p= 20 %	Percentile	Categorical scale
Finland	10	15	0.26	0.59	1.04	0.83	1.41	0.04	0.41	0	65.2	60
United States	12	23	1.52	1.00	1.25	1.00	1.69	0.25	0.69	1	100	100
Sweden	11.4	19	1.14	0.88	1.19	0.95	1.61	0.19	0.61	0	82.6	60
Japan	9.5	12	-0.06	0.49	0.99	0.79	1.34	-0.01	0.34	0	52.2	50
Korea, Rep. of	10.8	17	0.76	0.76	1.13	0.90	1.52	0.13	0.52	0	73.9	60
Netherlands	9.4	9	-0.12	0.47	0.98	0.78	1.32	-0.02	0.32	0	39.1	50
UK	9.4	9	-0.12	0.47	0.98	0.78	1.32	-0.02	0.32	0	39.1	50
Canada	11.6	20	1.27	0.92	1.21	0.97	1.63	0.21	0.63	1	87.0	80
Australia	10.9	18	0.83	0.78	1.14	0.91	1.54	0.14	0.54	0	78.3	60
Singapore	7.1	1	-1.58	0.00	0.74	0.59	1.00	-0.26	0.00	-1	4.3	0
Germany	10.2	16	0.38	0.63	1.06	0.85	1.44	0.06	0.44	0	69.6	60
Norway	11.9	22	1.46	0.98	1.24	0.99	1.68	0.24	0.68	1	95.7	100
Ireland	9.4	9	-0.12	0.47	0.98	0.78	1.32	-0.02	0.32	0	39.1	50
Belgium	9.3	8	-0.19	0.45	0.97	0.78	1.31	-0.03	0.31	0	34.8	40
New Zealand	11.7	21	1.33	0.94	1.22	0.98	1.65	0.22	0.65	1	91.3	80
Austria	8.4	6	-0.76	0.27	0.88	0.70	1.18	-0.12	0.18	0	26.1	40
France	7.9	5	-1.08	0.16	0.82	0.66	1.11	-0.18	0.11	0	21.7	40
Israel	9.6	14	0.00	0.51	1.00	0.80	1.35	0.00	0.35	0	60.9	50
Spain	7.3	4	-1.46	0.04	0.76	0.61	1.03	-0.24	0.03	-1	17.4	40
Italy	7.2	3	-1.52	0.02	0.75	0.60	1.01	-0.25	0.01	-1	13.0	20
Czech Republic	9.5	12	-0.06	0.49	0.99	0.79	1.34	-0.01	0.34	0	52.2	50
Hungary	9.1	7	-0.31	0.41	0.95	0.76	1.28	-0.05	0.28	0	30.4	40
Slovenia	7.1	1	-1.58	0.00	0.74	0.59	1.00	-0.26	0.00	-1	4.3	0

Sometimes, there is no need to normalise the indicators, if the indicators are already expressed with the same standard. See, for example, the case of the e-business readiness (Nardo *et al*., 2004), where all the indicators are expressed in terms of percentages of enterprises possessing a given infrastructure or using a given ICT tool. In such case the normalisation would rather obfuscate the issue, as one would loos the inherent information contained in the percentages.

WEIGHTING AND AGGREGATION

Weights based on statistical models

Principal components analysis, and more specifically factor analysis, group together sub-indicators that are collinear to form a composite indicator that captures as much of common information among sub-indicators as possible. Note that sub-indicators must have the same unit of measurement. Each factor (usually estimated using principal components analysis) reveals the set of indicators having the highest association with it. The idea under PCA/FA is to account for the highest possible variation in the indicators set using the smallest possible number of factors. Therefore, the composite no longer depends upon the dimensionality of the dataset but it is rather based on the "statistical" dimensions of the data.

According to PCA/FA, weighting only intervenes to correct for the overlapping information of two or more correlated indicators, and it is not a measure of theoretical importance of the associated indicator. If no correlation between indicators is found, then weights can not be obtained estimated with this method. This is the case for the new economic sentiment indicator, where factor and principal components analysis excluded the weighing of individual questions within a sub-component of the composite index (see the supplement B of the Business and Consumer Surveys Result N. 8/9 August/September 2001[14]). PCA/FA was excluded in the construction of an indicator of environmental sustainability when it was found that this procedure assigned negative weights to some sub-indicators (World Economic Forum, 2002).

The first step in FA is to check the correlation structure of the data, as explained in the section of multivariate analysis. If the correlation between the indicators is low, then it is unlikely that they share common factors. The second step is the identification of a certain number of latent factors, smaller than the number of sub-indicators, representing the data. Each factor depends on a set of coefficients (loadings), each coefficient measuring the correlation between the individual indicator and the latent factor. Principal component analysis is usually used to extract factors (Manly, 1994[15]). For a factor analysis only a subset of principal components are retained (*m*), the ones that account for the largest amount of the variance.

The standard practice is to choose factors that: (i) have associated eigenvalues larger than one; (ii) individually contribute to the explanation of overall variance by more than 10%; (iii) cumulatively contribute to the explanation of the overall variance by more than 60%. With the TAI reduced dataset (the one with 23 countries) the factors with eigenvalues close to the unity are the first four, as summarised in **Table 15**. Individually they explain more than 10% of the total variance and overall they count for about the 87% of variance.

Table 15. Eigenvalues of TAI dataset

Eigenvalues			
	Eigenval	% total Variance	Cumul. %
1	**3.3**	**41.9**	41.9
2	**1.7**	**21.8**	63.7
3	**1.0**	**12.3**	76.0
4	**0.9**	**11.1**	**87.2**
5	0.5	6.0	93.2
6	0.3	3.7	96.9
7	0.2	2.2	99.1
8	0.1	0.9	100.0

The third step deals with the rotation of factors (**Table 16**). The rotation (usually the *varimax rotation*) is used to minimise the number of sub-indicators that have a high loading on the same factor. The idea in transforming the factorial axes is to obtain a "simpler structure" of the factors (ideally a structure in which each indicator is loaded exclusively on one of the retained factors). Rotation is a standard step in factor analysis, it changes the factor loadings and hence the interpretation of the factors leaving unchanged the analytical solutions obtained *ex-ante* and *ex-post* the rotation.

Table 16. Factor loadings of TAI based on principal components

	Factor 1	Factor 2	Factor 3	Factor 4	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*
Patents	0.07	0.97	0.06	0.06	*0.00*	***0.68***	*0.00*	*0.00*
Royalties	0.13	0.07	-0.07	0.93	*0.01*	*0.00*	*0.00*	***0.49***
Internet	0.79	-0.21	0.21	0.42	***0.24***	*0.03*	*0.04*	*0.10*
Tech exports	-0.64	0.56	-0.04	0.36	*0.16*	***0.23***	*0.00*	*0.07*
Telephones	0.37	0.17	0.38	0.68	*0.05*	*0.02*	*0.12*	***0.26***
Electricity	0.82	-0.04	0.25	0.35	***0.25***	*0.00*	*0.05*	*0.07*
Schooling	0.88	0.23	-0.09	0.09	***0.29***	*0.04*	*0.01*	*0.00*
University	0.08	0.04	0.96	0.04	*0.00*	*0.00*	***0.77***	*0.00*
Expl.Var	2.64	1.39	1.19	1.76				
Expl./Tot	*0.36*	*0.26*	*0.24*	*0.42*				

The last step deals with the construction of the weights from the matrix of factor loadings after rotation, given that the square of factor loadings represent the proportion of the total unit variance of the indicator which is explained by the factor. The approach used by Nicoletti G., Scarpetta S., Boylaud O. (2000) is that of grouping the sub-indicators with the highest factors loadings in *intermediate* composite indicators. With the TAI dataset the *intermediate* composites are 4 (Table 6.2). The first includes Internet (with a weight of 0.24), Electricity (weight 0.25) and Schooling (weight 0.29).[16] Likewise the second *intermediate* is formed by Patents and Technology Exports (worth 0.68 and 0.23 respectively), the third only by University (0.77) and the fourth by Royalties and Telephones (weighted with 0.49 and 0.26).

Then the four *intermediate* composites are aggregated by weighting each composite using the proportion of the explained variance in the dataset: 0.36 for the first (0.36 = 2.64/(2.64+1.39+1.19+1.76)), 0.26 for the second, 0.24 for the third and 0.42 for the fourth.[17] Notice that different methods for the extraction of principal components imply different weights, hence different scores for the composite (and possibly different country ranking). For example if Maximum Likelihood (ML) were to be used instead of Principal Component (PC) the weights obtained would be as given in **Table 17**.

Table 17. Factor loadings of TAI based on maximum likelihood

	M L	PCA
Patents	0.19	0.17
Royalties	0.20	0.20
Internet	0.07	0.08
Tech exports	0.07	0.06
Telephones	0.15	0.11
Electricity	0.11	0.09
Schooling	0.19	0.10
University	0.02	0.18

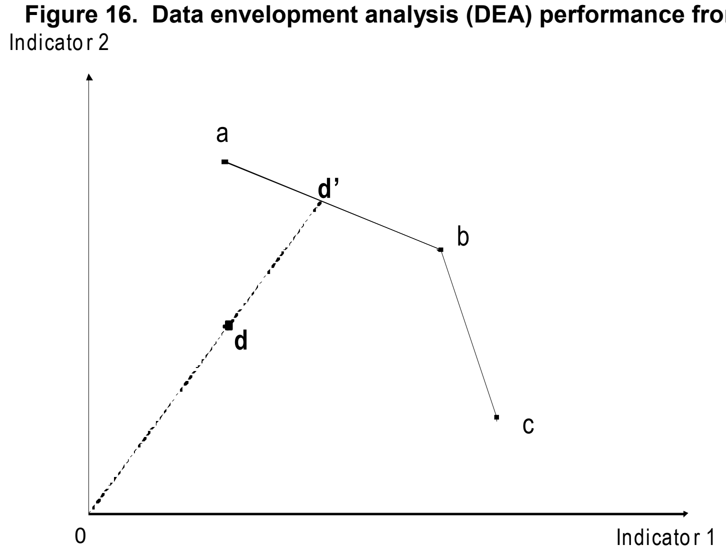
Data envelopment analysis (DEA)

Data Envelopment Analysis (DEA) employs linear programming tools to estimate an *efficiency frontier* that would be used as a benchmark to measure the relative performance of countries.[18] This requires construction of a benchmark (the *frontier*) and the measurement of the distance between countries in a multi-dimensional framework. The following assumptions are made for the benchmark:

(i) positive weights--the higher the value of a given sub-indicator, the better for the corresponding country;
(ii) non-discrimination of countries that are the best in any single dimension (sub-indicator), thus ranking them equally; and
(iii) a linear combination of the best performers is feasible, i.e., convexity of the frontier.

The distance of each country with respect to the benchmark is determined by the location of the country and its position relative to the frontier. Both issues are represented in **Figure 16**, for the simple case of four countries and two base indicators that are represented in the two axes. Countries (*a, b, c, d*) are ranked according to the score of the indicators. The line connecting countries *a*, *b* and *c* constitutes the performance frontier and the benchmark for country *d* which lies beyond the frontier. The countries supporting the frontier are classified as the *best performing*, while country *d* is the *worst performing*.

Figure 16. Data envelopment analysis (DEA) performance frontier



Source: Rearranged from Mahlberg and Obersteiner (2001).

The performance indicator is the ratio of the distance between the origin and the actual observed point and that of the projected point in the frontier: $\overline{0d} / \overline{0d'}$. The best performing countries will have a performance score of 1, and the least performing less than one. This ratio corresponds to $(w_{1d}I_{1d} + w_{2d}I_{2d})/(w_{1d}I^*_{1d} + w_{2d}I^*_{2d})$, where I^*_{id} is the frontier value of indicator *i*, *i=1,2*, and I_{id} is its actual value (see expression 6.1 for more than 2 indicators). The set of weights for each country therefore depends on its position with respect to the frontier, while the benchmark corresponds to the ideal point with a similar mix of indicators (*d'* in the example). The benchmark could also be determined by a hypothetical decision-maker (Korhonen et al. 2001), who locates the target in the efficiency frontier with the most preferred combination of sub-indicators. This is similar to the budget allocation method (see below) where experts are asked to assign weights (i.e. priorities) to sub-indicators.

Benefit of the doubt approach (BOD)

The application of the DEA to the field of composite indicators is known as the benefit of the doubt approach (BOD) and it was originally proposed to evaluate macroeconomic performance (Melyn and Moesen, 1991)[19]. In the BOD approach, the composite indicator is defined as the ratio of a country's actual performance over its benchmark performance:

$$CI_c = \frac{\sum_{c=1}^{M} I_{qc} w_{qc}}{\sum_{c=1}^{M} I^*_{qc} w_{qc}} \tag{18}$$

where I_{qc} is the normalised (with the max-min method) score of qth sub-indicator (q=1,…,Q) for country c (c=1,…M) and wqc the corresponding weight. Cherchye et al. (2004) who firstly implemented this method suggested obtaining the benchmark as solution of a maximisation problem, although external benchmarks are also possible:

$$I^* = I^*(w\ ) = \underset{I_k, k\in\{1,\ldots,M\}}{arg\,max}\left(\sum_{q=1}^{Q} I_{qk} w_q\right) \tag{19}$$

I* is the score of the hypothetical country that maximises the overall performance (defined as the weighted average), given the (unknown) set of weights w. Note that (i) weights are country specific: different sets of weights may lead to choose different countries as far as there is no country having the highest score in all sub-indicators; (ii) the benchmark would in general be country-dependent, so no unique benchmark would exist (unless, as before, a country is better in all sub-indicators), (iii) sub-indicators must be comparable, i.e. have the same unit of measurement.

The second step is the specification of the set of weights for each country. The optimal set of weighs--if it exists--guarantees the best position for the associated country *vis-à-vis* all other countries in the sample. With any other weighting profile, the relative position of that country would have been worse. Optimal weights are obtained by solving the following constrained optimisation:

$$CI^*_c = \underset{w_{qc}, q=1,\ldots,Q}{arg\,max} \frac{\sum_{q=1}^{Q} I_{qc} w_{qc}}{\underset{I_k, k\in\{1,\ldots,M\}}{max}\left(\sum_{q=1}^{Q} I_{qk} w_{qc}\right)} \quad \text{for } c=1,..,M \tag{20}$$

subject to non negativity constraints on weights.[20]

The resulting composite index will range between zero (lowest possible performance) and 1 (the benchmark). Operationally, equation (20) can be reduced to a linear programming problem (21) by multiplying all the weights with a common factor that does not alter the index value and then solved using optimisation algorithms

$$
\begin{aligned}
& CI_c^* = \underset{wqc}{arg\,max} \sum_{q=1}^{Q} I_{qc} w_{qc} \\
& s.t. \\
& \sum_{q=1}^{Q} I_{qk} w_{qk} \leq 1 \\
& w_{qk} \geq 0 \\
& \forall k = 1,...,M; \forall q = 1,...,Q
\end{aligned}
\qquad (21)
$$

The result of BOD approach applied to the TAI example can be seen in **Table 19**. Weights are given in the first eight columns, while the last column contains the composite indicator values. Finland, the United States and Sweden have a composite index value of one, i.e. they have the top score in the ranking. This however hides a problem of multiple equilibria. In Figure 18, any point between country *a* (e.g., Finland) and country *b* (e.g., USA) can be an optimal solution for these countries. Thus weights are not uniquely determined. Notice also that the multiplicity of solutions is likely to depend upon the set of constraints imposed to the weights of the maximisation problem in (21) – the wider is the range of variation of weights, the lower is the possibility of obtaining a unique solution[21]. Second, the set of weights for each country as calculated by the above algorithm[22] does not sum up to one, making the comparison between countries and with other methods (like FA or EW) impossible.

Table 19. Benefit of the doubt (BOD) approach applied to TAI

	Patents	Royalties	Internet	Tech. Export	Telephones	Electricity	Schooling	University	CI
Finland	0.15	0.17	0.17	0.16	0.19	0.17	0.17	0.19	**1**
United States	0.20	0.20	0.17	0.21	0.15	0.15	0.21	0.14	**1**
Sweden	0.18	0.21	0.15	0.19	0.19	0.16	0.20	0.14	**1**
Japan	0.22	0.15	0.15	0.22	0.22	0.16	0.21	0.15	**0.87**
Korea	0.22	0.14	0.14	0.22	0.14	0.14	0.22	0.22	**0.80**
Netherlands	0.22	0.22	0.14	0.22	0.22	0.14	0.14	0.14	**0.75**
United Kingdom	0.14	0.21	0.14	0.21	0.21	0.14	0.20	0.15	**0.71**
Canada	0.14	0.14	0.14	0.21	0.21	0.21	0.21	0.14	**0.73**
Australia	0.13	0.13	0.20	0.13	0.13	0.20	0.20	0.20	**0.66**
Singapore	0.14	0.14	0.14	0.20	0.20	0.20	0.14	0.20	**0.62**
Germany	0.22	0.15	0.15	0.22	0.21	0.15	0.22	0.15	**0.62**
Norway	0.14	0.14	0.20	0.14	0.20	0.20	0.20	0.14	**0.86**
Ireland	0.14	0.21	0.14	0.21	0.21	0.14	0.20	0.15	**0.60**
Belgium	0.14	0.16	0.14	0.21	0.19	0.21	0.21	0.14	**0.54**
New Zealand	0.21	0.14	0.21	0.14	0.14	0.21	0.21	0.14	**0.58**
Austria	0.22	0.14	0.14	0.22	0.22	0.22	0.14	0.14	**0.52**
France	0.22	0.14	0.14	0.22	0.22	0.22	0.14	0.14	**0.51**
Israel	0.21	0.15	0.15	0.22	0.22	0.15	0.22	0.15	**0.49**
Spain	0.21	0.14	0.14	0.21	0.21	0.14	0.14	0.21	**0.34**
Italy	0.22	0.14	0.14	0.22	0.22	0.22	0.14	0.14	**0.38**
Czech Rep.	0.22	0.15	0.15	0.22	0.15	0.22	0.22	0.15	**0.31**
Hungary	0.22	0.14	0.21	0.22	0.14	0.14	0.22	0.15	**0.27**
Slovenia	0.22	0.14	0.14	0.22	0.22	0.22	0.14	0.14	**0.28**

Note: Columns 1 to 8 : weights, column 9: composite indicator for a given country, n=23 countries.

Unobserved components model (UCM)

In the unobserved components model (UCM), sub-indicators are assumed to depend on an unobserved variable plus an error term, e.g. the "percentage of firms using internet in country *j*" depends upon the (unknown) propensity to adopt new information and communication technologies plus an error term, accounting, for example, for the error in the sampling of firms. Therefore, estimating the unknown component sheds some light on the relationship between the composite and its components. The weight obtained will be set to minimise the error in the composite. This method resembles the well known regression analysis. The main difference resides in the dependent variable, which is unknown under UCM.

Let *ph(c)* be the unknown phenomenon to be measured. The observed data consist on a cluster of *q=1,…,Q(c)* indicators, each measuring an aspect of *ph(c)*. Let *c=1,…M(q)* the countries covered by indicator *q*. The observed score of country *c* on indicator *q*, *I(c,q)*, can be written as a linear function of the unobserved phenomenon and an error term, $\varepsilon(c,q)$:

$$I(c,q) = \alpha(q) + \beta(q)[ph(c) + \varepsilon(c,q)] \tag{22}$$

$\alpha(q)$ and $\beta(q)$ are unknown parameters mapping *ph(c)* on *I(c,q)*.

The error term captures two sources of uncertainty. First, the phenomenon can be imperfectly measured or observed in each country (e.g. errors of measurement). Second, the relationship between *ph(c)* and *I(c,q)* can be imperfect (e.g. *I(c,q)* may only be a noisy indicator of the phenomenon, if there are differences among countries about the indicator). The error term $\varepsilon(c,q)$ is assumed to have a zero mean, $E(\varepsilon(c,q)) = 0$, and the same variance across countries within a given indicator, but a different variance across indicators, $E(\varepsilon(c,q)^2) = \sigma_q^2$; it also holds $E(\varepsilon(c,q)\varepsilon(i,h)) = 0$ for $c \neq i$ or $q \neq h$.

The error term is assumed to be independent across indicators, given each sub-indicator should ideally measure a particular aspect of the phenomenon independent of others. Furthermore, it is usually assumed that *ph(c)* is a random variable with mean zero and unit variance, and the indicators are re-scaled to take values between zero and one. The assumption that both *ph(c)* and $\varepsilon(c,q)$ are jointly normally distributed simplifies the estimation of the level of *ph(c)* in country *c*. This is done by using the mean of the conditional distribution of the unobserved component, once the observed scores are appropriately re-scaled:

$$E[ph(c)/I(c,1),\dots,I(c,Q(c))] = \sum_{q=1}^{Q(c)} w(c,q)\frac{I(c,q)-\alpha(q)}{\beta(q)} \tag{23}$$

The weights are equal to:

$$w(c,q) = \frac{\sigma_q^{-2}}{1+\sum_{q=1}^{Q(c)}\sigma_q^{-2}} \tag{24}$$

w(c,q) is a decreasing function of the variance of indicator *q*, and an increasing function of the variance of the other indicators. The weight, *w(c,q),* depends on the variance of indicator *q* (numerator) and on the sum of the variances of the all the other sub-indicators, including *q* (denominator). However, since not all countries have data on all sub-indicators, the denominator of *w(c,q)* could be country-specific. This may produce non-comparability of country values for the composite as in BOD. Obviously whenever the set of indicators is equal for all countries, weights will be no longer country specific and comparability will be assured. The variance of the conditional distribution is given by:

$$var[\,ph(c)/I(c,1),...,I(c,Q(c))] = [\,1+\sum_{q=1}^{Q(c)} \sigma_q^{-2}\,]^{-1} \qquad (25)$$

and can be used as a measure of the precision of the composite. The variance decreases in the number of indicators for each country, and increases in the variance of the disturbance term for each indicator. The estimation of the model could be simplified by the assumption of normality for *ph(c)* and $\varepsilon(c,q)$. The likelihood function of the observed data is maximised with respect to the unknown parameters, $\alpha(q)s$, $\beta(q)s$, and $\sigma_q^2 s$, and their estimated values substituted in equation (23) to obtain the composite indicator and the weights.

Budget allocation (BAL)

It is essential to bring together experts that have a wide spectrum of knowledge, and experience to ensure that a proper weighting system is established. Special care should be given in the identification of the population of experts from which to draw a sample, stratified or otherwise[23]. The budget allocation method (BAL) has four different phases:

- Selection of experts for the valuation;
- Allocation of budget to the sub-indicators;
- Calculation of the weights;
- Iteration of the budget allocation until convergence is reached (optional).

Public opinion

From a methodological point of view, opinion polls focus on the notion of “concern”. That is people are asked to express the degree of concern (e.g., much or little) on issues, measured by the base indicators. As with expert assessments, the budget allocation method could also be applied in public opinion polls. However it is more difficult to ask the public to allocate a hundred points to several sub-indicators than to express a degree of concern about a given problem.

Analytic hierarchy process (AHP)

The Analytic Hierarchy Process (AHP) is a widely used technique for multi-attribute decision making (Saaty, 1987). It enables the decomposition of a problem into hierarchy and assures that both qualitative and quantitative aspects of a problem are incorporated in the evaluation process, during which opinions are systematically extracted by means of pairwise comparisons. According to Forman et al. (1983): “*AHP is a compensatory decision methodology because alternatives that are efficient with respect to one or more objectives can compensate by their performance with respect to other objectives. AHP allows for the application of data, experience, insight, and intuition in a logical and thorough way within a hierarchy as a whole. In particular, AHP as weighting method enables decision-maker to derive weights as opposed to arbitrarily assign them.*”

Weights represent the trade-off across indicators. They measure the willingness to forego a given variable in exchange for another. Hence, they are not importance coefficients. This could create a misunderstanding, if AHP weights are interpreted as importance coefficients (see Ülengin et al. 2001).

The core of AHP is an ordinal pair-wise comparison of attributes. For a given objective, the comparisons are made per pairs of sub-indicators: Which of the two is the more important? And, by how much? The preference is expressed on a semantic scale of 1 to 9. A preference of 1 indicates equality between two sub-indicators, while a preference of 9 indicates that the sub-indicator is 9 times more important than the other one. The results are represented in a comparison matrix (**Table 20**), where $A_{ii} = 1$ and $A_{ij} = 1 / A_{ji}$.

Table 20. Comparison matrix of eight TAI sub-indicators

Objective	Patents	Royalties	Internet	Tech exports	Telephon e	Electricity	Schooling	University
Patents	1	2	3	2	5	5	1	3
Royalties	1/2	1	2	1/2	4	4	½	3
Internet	1/3	½	1	1/4	2	2	1/5	1/2
Tech. exports	1/2	2	4	1	4	4	1/2	3
Telephones	1/5	1/4	1/2	1/4	1	1	1/5	1/2
Electricity	1/5	¼	1/2	1/4	1	1	1/5	1/2
Schooling	1	2	5	2	5	5	1	4
University	1/3	1/3	2	1/3	2	2	1/4	1

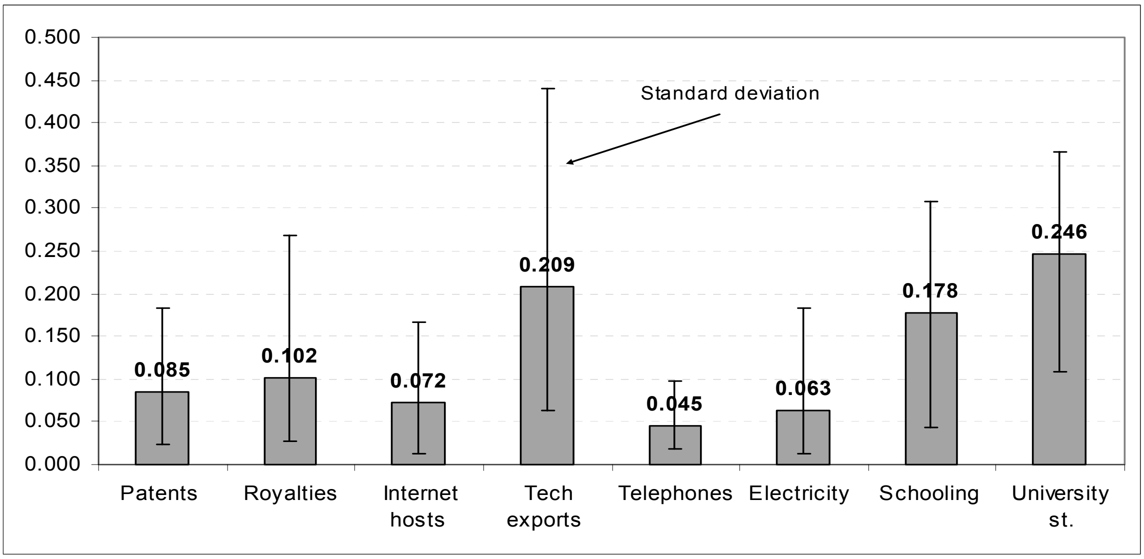
For the example, *patents* is three times more important than *Internet*. Each judgement reflects the perception of the relative contributions (weights) of the two sub-indicators to the overall objective (**Table 21**).

Table 21. Comparison matrix of three TAI sub-indicators

Objective	Patents	Royalties	Internet
Patents	w_P/w_P	w_P/w_{ROY}	w_P/w_I
Royalties	w_{ROY}/w_P	w_{ROY}/w_{ROY}	w_{ROY}/w_I
Internet	w_I/w_P	w_I/w_{ROY}	w_I/w_I

The relative weights of the sub-indicators are calculated using an eigenvector. This method enables to check the consistency of the comparison matrix through the calculation of the eigenvalues. **Figure 17** shows the results of the evaluation process and the weights, together with the corresponding standard deviation[24].

Figure 17. Analytic hierarchy process (AHP) weighting of TAI



Note. Average weight (bold) and standard deviation.

People's beliefs however are not always consistent. For example, if one claims that A is much more important than B, B slightly more important than C, and C slightly more important than A, his/her judgement is inconsistent and the results are less trustworthy. Inconsistency, however, is part of the human nature. Therefore it could be enough to measure the degree of inconsistency, such that results could be acceptable in the public eye. For a matrix of size Q × Q, only Q–1 comparisons are required to establish weights for Q indicators. The actual number of comparisons performed in AHP is Q(Q–1)/2. This is computationally costly, but results in a set of weights that is less sensitive to errors of judgement. In addition, the redundancy allows for a measure of judgement errors, an inconsistency ratio. Small inconsistency ratios--the suggested rule-of-thumb is less than 0.1, although 0.2 is often cited--do no drastically affect the weights (Saaty, 1980; Karlsson, 1998).

Conjoint analysis (CA)

Merely asking respondents how much importance they attach to a sub-indicator is unlikely to yield effective “willingness to pay” valuations. These can be inferred by using conjoint analysis (CA) from respondents’ ranking of alternative scenarios (Hair et al. 1995). The conjoint analysis is a decompositional multivariate data analysis technique frequently used in marketing (McDaniel and Gates, 1998) and consumer research (Green and Srinivasan, 1978). If AHP derives the “worth” of an alternative, *summing up* the “worth” of the individual sub-indicators, the CA does the opposite, i.e. it disaggregates preferences.

This method asks for an evaluation (a preference) over a set of alternative scenarios. A scenario can be a given set of values for the sub-indicators. The preference is then decomposed by relating the single components (the known values of sub-indicators of that scenario) to the evaluation. Although this methodology uses statistical analysis to treat data, it relies on the opinion of people, *e.g.* experts, politicians, citizens, who are asked to choose which set of sub-indicators they prefer, with each person presented with different choice sets to evaluate.

The absolute value (or level) of sub-indicators could be varied both within the choice sets presented to the same individual and across individuals. A preference function would be estimated using the information coming from the different scenarios. Therefore a probability of the preference could be estimated as a function of the levels of the sub-indicators defining the alternative scenarios:

$$pref_c = P(I_{1c}, I_{2c}, \ldots, I_{Qc}) \tag{26}$$

where I_{qc} is the level of sub-indicator $q=1,\ldots,Q,$ for country $c=1,\ldots,M$. After estimating this probability (often using discrete choice models), the derivatives with respect to the sub-indicators of the preference function can be used as weights to aggregate the sub-indicators in a composite index:

$$CI_c = \sum_{q=1}^{Q} \frac{\partial P}{\partial I_{qc}} I_{qc} \tag{27}$$

The idea is to calculate the total differential of the function P at the point of indifference between alternative states of nature. Solving for the sub-indicator $q,$ one obtains the marginal rate of substitution of I_{qc}. Therefore $\partial P / \partial I_{qc}$ (thus the weight) indicates a trade-off--how the preference changes with the change of the indicator. This implies compensability among indicators, i.e. the possibility of offsetting the lack in some dimension with an outstanding performance in another dimension. This is an important feature of this method, and should be carefully evaluated vis-à-vis the objectives of the whole analysis. For example, compensability might not be desirable when dealing with environmental issues.

Performance of the different weighting methods

The weights for the TAI example are calculated using different weighting methods -- equal weighting (EW), factor analysis (FA), budget allocation (BAL), and analytical hierarchy process (AHP) (**Table 22**). The diversity in the resulting weights from applying different methods is noteworthy. Clearly with each method, different sub-indicators are evaluated in a different way. Patents, for example, are worth 17% of the weight according to the FA, but only 9% according to the AHP. This deeply influences the variability of each country's ranking (**Table 23**). For example, Korea ranks second with the AHP, but only fifth, when EW or FA are used. AHP assigns high weights (more than 20%) to two indicators, *High tech exports* and *University enrolment ratio*, for which Korea has higher scores for one or both indicators, compared to the United States, Sweden or Japan.The role of the variability in the weights and their influence in the value of the composite are discussed in the section on sensitivity analysis.

Table 22. TAI weights based on different methods

Equal weighting (EW), factor analysis (FA), budget allocation (BAL), and analytic hierarchy process (AHP)

Methods/ Sub-indicators	Patents	Royalties	Internet	Tech exports	Telephones	Electricity	Schooling	University
EW	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
FA	0.17	0.15	0.11	0.06	0.08	0.13	0.13	0.17
BAL	0.11	0.11	0.11	0.18	0.10	0.06	0.15	0.18
AHP	0.09	0.10	0.07	0.21	0.05	0.06	0.18	0.25

Table 23. TAI country rankings based on different weighting methods

Weighting method/Country	EW	FA	BOD	BAL	AHP
Finland	1	1	1	1	1
United States	2	2	1	2	3
Sweden	3	3	1	3	4
Japan	4	4	4	5	5
Korea, Rep. of	5	5	6	4	2
Netherlands	6	6	7	8	11
United Kingdom	7	8	9	7	7
Singapore	8	11	12	6	6
Canada	9	10	8	10	10
Australia	10	7	10	11	9
Germany	11	12	11	9	8
Norway	12	9	5	13	16
Ireland	13	14	13	12	12
Belgium	14	15	15	14	13
New Zealand	15	13	14	17	18
Austria	16	16	16	15	15
France	17	17	17	16	14
Israel	18	18	18	18	17
Spain	19	19	20	19	19
Italy	20	20	19	21	21
Czech Republic	21	21	21	22	22
Hungary	22	23	23	20	20
Slovenia	23	22	22	23	23

Note: e.g., the United States ranks first according to BOD, second according to EW, FA, and BAL and third according to AHP.

Table 24. Advantages and disadvantages of different weighting methods

Advantages	Disadvantages
Benefit of the doubt (BOD) -- *e.g.*, Human Development Index (Mahlberg and Obersteiner, 2001); Sustainable Development (Cherchye and Kuosmanen, 2002); Social Inclusion (Cherchye, Moesen, Van Puyenbroeck, 2004); Macro-economic performance evaluation (Melyn and Moesen, 1991, and Cherchye 2001); Unemployment (Storrie and Bjurek, 1999, and 2000).	
The indicator is sensible to national policy priorities, in that the weights are endogenously determined by the observed performances (this is a useful second best approach whenever the first best – full information about true policy priorities- can not be attained). The benchmark is not based upon theoretical bounds, but a linear combination of observed best performances. It is useful in policy arena, since policy makers could not complain about unfair weighting: any other weighting scheme would have generated lower composite scores. Such an index could be "incentive generating" rather than "punishing" the countries lagging behind. Weights, by revealing information about the policy priorities, may help to define trade-offs, overcoming the difficulties of linear aggregations.	Weights are country specific, thus cross-country comparisons is not possible. Without imposing constraints on weights (except the non-negativity) the most likely solution is to have all countries with a composite equal to 1. It may happen that there exist a multiplicity of solutions making the optimal set of weights undetermined (this is likely to happen when the CI=1). Different normalisation of the scores is likely to give different weighting schemes. The index is likely to reward the status-quo, since for each country the maximisation problem gives higher weights to higher scores. Endogenous weighting has the risk of substituting open experts' opinions with the analyst's manipulation of weights (through the constraints). Transparency of the procedure would be lost. The value of the scoreboard depends on the benchmark performance. If this changes the composite will change as well as the set of weights (and the country ranking). Moreover if one country is "lazy" in all but one variable where it is the best performing, then it could be a benchmark in the frontier The best performer (the one with a composite equal to one) will not see its progress reflected in the composite (that will remain stacked to 1). This can be solved by imposing an external benchmark.
Unobserved Components Models -- *e.g.*, Governance indicators (see Kaufmann, Kraay and Zoid-lobatón, 1999 and 2003)	
Weights do not depend on ad hoc restrictions. It can be used even if component indicators are not correlated.	Reliability and robustness of results depend on the availability of enough data. With highly correlated sub-indicators there could be identification problems. Thus the method is likely to work well with independent sub-indicators. The method rewards the absence of outliers, given that weights are a decreasing function of the variance of sub-indicators. If each country has a different number of sub-indicators; weights are hence country specific
Budget Allocation -- *e.g.*, Employment Outlook (OECD,1999); Composite Indicator on e-Business Readiness (EC-JRC, 2004b);National Health Care System Performance (King's Fund., 2001);Eco-indicator 99 (Pré-Consultants NL, 2000) (weights based on survey from experts); Overall Health System Attainment (WHO, 2000) (weights based on survey from experts).	
Weighting is based on experts' opinion and not on technical manipulations. Experts' opinions are likely to increase the legitimacy of the composite and create a forum of discussion around which to form a consensus for policy action.	Weighting reliability. Weights could reflect specific local conditions (e.g. in environmental problems), so expert weighting may not be transferable from one area to another. Allocating a certain budget over a too large number of indicators can give serious cognitive stress to the experts, as it implies circular thinking. The method is likely to produce inconsistencies for a number of indicators higher than 10. The weighting may not measure the importance of each sub-indicator but rather the urgency or need for political intervention in the dimension of the sub-indicator concerned (e.g. more weight on Ozone emissions if the expert feels that not enough has been made to abate them).
Public Opinion – *e.g.* **c**oncern about environmental problems Index (Parker, 1991)	

Deals with issues on the public agenda. Allows all stakeholders to express their preference, and creates a consensus for policy actions.	Implies the measurement of "concern" . The method could produce inconsistencies when dealing with high number of indicators.
Analytic Hierarchy Process -- *e.g.*, Index of Environmental Friendliness, (Puolamaa *et al.*, 1996).	
The method can be used both for qualitative and quantitative data. The transparency of the composite is higher	The method requires a high number of pairwise comparisons and thus it can be computationally costly. The results depends on the set of evaluators chosen and the setting of the experiment
Conjoint Analysis – *e.g.*, indicator of quality of life in the city of Istanbul (Ülengin et al. 2001); advocated by Kahn (1998) and Kahn and Maynard (1996) for environmental applications.	
Weights represent trade-offs across indicators It takes into account the socio-political context, and the values of respondents.	It needs a pre-specified utility function and it implies compensability. Depends on the sample of respondents chosen and on how questions are framed. It requires a large sample of respondents and each respondent may be required to express a large number of preferences. The estimation process is complex.

Additive aggregation methods

The simplest additive aggregation method entails the calculation of the ranking of each country according to each sub-indicator and summation of the resulting ranking, e.g., Information and Communication Technologies Index (Fagerberg, 2001). The method is based on ordinal information (the so called Borda rule). It is simple and independent of outliers. But, the absolute value of information is lost.

$$CI_c = \sum_{q=1}^{Q} Rank_{qc} \text{ for } c=1,\ldots,M. \tag{28}$$

The second method is based on the number of indicators that are above and below some benchmark. This method uses nominal scores for each indicator to calculate the difference between the number of indicators that are above and below an arbitrarily defined threshold around the mean, e.g., the Innovation Scoreboard (European Commission, 2001a).

$$CI_c = \sum_{q=1}^{Q} \cdot sgn\left[\frac{I_{qc}}{I_{EUq}} - (1+p)\right] \text{ for } c=1,\ldots,M. \tag{29}$$

The threshold value p can be arbitrarily chosen above or below the mean. As with the preceding method, it is simple and unaffected by outliers. However, the interval level information is lost. For example, assume that the value of indicator I for country a is 30% above the mean and the value for country b is 25% above the mean, with a threshold of 20% above the mean. Both country a and b are then counted equally as 'above average', in spite of a having a higher score than b.

By far, the most widespread linear aggregation is the summation of weighted and normalised sub-indicators:

$$CI_c = \sum_{q=1}^{Q} w_q I_{qc} \tag{30}$$

with $\sum_q w_q = 1$ and $0 \leq w_q \leq 1$, for all $q=1,..,Q$ and $c=1,...,M$.

Although widely used, this aggregation imposes restrictions on the nature of sub-indicators. In particular obtaining a meaningful composite indicator depends on the quality of the underlying sub-indicators and the unit of measurement of these sub-indexes. Furthermore, additive aggregations have important implications on the interpretation of weights.

When using a linear additive aggregation technique, a necessary and sufficient condition for the existence of a proper composite indicator is *preference independence*: given the sub-indicators $\{x_1, x_2, ..., x_Q\}$, an additive aggregation function exists if and only if these indicators are mutually preferentially independent[25] (Debreu, 1960; Keeney and Raiffa, 1976; Krantz et al., 1971).

Preferential independence is a very strong condition, as it implies that the trade-off ratio between two variables $S_{x,y}$ is independent of the values of the Q-2 other variables, (Ting, 1971)[26]. From an operational point of view, this means that an additive aggregation function permits the assessment of the marginal contribution of each variable separately. These marginal contributions can then be added together to yield a total value. If, for example, environmental dimensions are involved, the use of a linear aggregation procedure implies that among the different aspects of an ecosystem, there are no synergies or conflict. This appears to be quite an unrealistic assumption (Funtowicz et al., 1990). For example, *"laboratory experiments made clear that the combined impact of the acidifying substances SO_2, NO_X, NH_3 and O_3 on plant growth is substantially more severe that the (linear) addition of the impacts of each of these substances alone would be."* (Dietz and van der Straaten, 1992). Additive aggregation thus could result in a biased composite indicator, i.e. it will not entirely reflect the information of its sub-indicators. The dimension and the direction of the error are not easily determined, and the composite can not be adjusted properly.

Non-compensatory multicriteria approach (MCA)

As a common practice, greater weight could be given to components which are considered to be more significant in the context of the particular composite indicator, (OECD, 2003). Yet, it can be shown that when using an additive or a multiplicative aggregation rule and sub-indicators are expressed as intensities (e.g. in pounds, litres or euro and not qualities – e.g. good, bad, medium – or in rankings) the substitution rates equal the weights of the variables up to a multiplicative coefficient[27] (Munda and Nardo 2003). As a consequence, weights in additive aggregations necessarily have the meaning of substitution rates (*trade-offs*) and do not indicate the importance of the indicator associated. This implies a compensatory logic, i.e. the possibility of offsetting a disadvantage on some variables by a sufficiently large advantage on other variables. For example, in the construction of the TAI index a compensatory logic (using equal weighting) would imply that one is willing to renounce, let's say, to 2% of *Patents granted to residents,* or to 2% of *University enrolment* in exchange of a 2% increase in *Electricity consumption*.

The implication is the existence of a theoretical inconsistency in the way weights are actually used and their real theoretical meaning. For the weights to be interpreted as "*importance coefficients*" (the greatest weight is placed beside the most important "dimension") non-compensatory aggregation procedures must be used to construct composite indicators (Podinovskii, 1994). This can be done using a non-compensatory multi-criteria approach.

When various variables are used to evaluate a set of countries, some of these variables may be in favour of one country while other variables may be in favour of another. As a consequence a conflict among the variables could arise. This conflict can be treated in the light of a non-compensatory logic by taking into account the absence of preferential independence within a discrete non-compensatory multi-criteria approach (NCMC) (Munda, 1995; Roy, 1996; Vincke, 1992).

Given a set of sub-indicators $\boldsymbol{G}=\{x_q\},\ q=1,\dots,Q,$ and a finite set $\boldsymbol{M}=\{c\},\ c=1,\dots,M$ of countries, assume that the evaluation of each country c with respect to an individual indicator x_q (i.e. the indicator score or variable) is based on an *interval or ratio* scale of measurement. For simplicity of exposition, it is also assumed that a higher value of an individual indicator is preferred to a lower one, i.e., the higher, the better. Further assume, the existence of a set of weights $w=\{w_q\},\ q=1,2,\dots Q,$ with $\sum_{q=1}^{Q} w_q = 1$, interpreted as *importance coefficients*. This information constitutes the impact matrix. For explanatory purposes, suppose only 5 of the countries included in the TAI dataset[28] and equal weights are given to all of the sub-indicators (**Table 25**).

Table 25. Impact matrix for TAI (five countries)

	Patents	Royalties	Internet	Tech exports	Telephones	Electricit y	Schooling	University
Finland	187	125.6	200.2	50.7	3.080	4.150	10	27.4
USA	289	130	179.1	66.2	2.997	4.073	12	13.9
Sweden	271	156.6	125.8	59.7	3.096	4.145	11.4	15.3
Japan	994	64.6	49	80.8	3.003	3.865	9.5	10
Korea	779	9.8	4.8	66.7	2.972	3.653	10.8	23.2
weight	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

The mathematical problem is then how to use this information to rank all the countries from the best to the worst one in a complete pre-order (i.e. without any incomparability relation, see Roubens and Vincke, 1985). The following are important:

- Intensity of preference (how much country a is better than country b according to sub-indicator q);
- Number of indicators in favour of a given country;
- Weight attached to each indicator;
- Relationship of each country with respect to all the others.

The sources of uncertainty and imprecise assessment should be reduced as much as possible. Unfortunately Arrow's impossibility theorem (Arrow, 1963) clearly shows that no perfect aggregation convention can exist. Therefore, when aggregating, it is essential to check not only which properties are respected by a given ranking procedure, but also that the essential properties the specific problem are not lost.

The mathematical aggregation convention can be divided into two main steps:

- Pair-wise comparison of countries according to the whole set of sub-indicators used.
- Ranking of countries in a complete pre-order.

The first step results in a $M \times M$ matrix, E, called *outranking matrix* (Arrow and Raynaud, 1986, Roy, 1996). Any generic element of E: e_{jk} , $j \neq k$ is the result of the pair-wise comparison, according to all the Q sub-indicators, between countries j and k. Such a global pair-wise comparison is obtained by means of equation:

$$e_{jk} = \sum_{q=1}^{Q} (w_q (Pr_{jk}) + \frac{1}{2} w_q (In_{jk})) \tag{31}$$

where $w_q(Pr_{jk})$ and $w_q(In_{jk})$ are the weights of sub-indicators presenting a preference and an indifference relation respectively. In other words, the score of country j is the sum of the weights of sub-indicators, for which this country does better than country *I,* as well as – if any – half of the weights for the sub-indicators according to which the two countries do equally well (**31)**. $e_{jk} + e_{kj} = 1$ clearly holds.

The pair-wise comparisons are different from those in the AHP method – which belongs to the set of compensatory multicriteria methods, jointly with CA. In the latter, the question to be answered was whether I_q is more important than I_z, here, instead, the question is whether I_q is higher for country *a* or for country *b*. And if I_q is indeed higher for country *a* , it is the weight of sub-indicator *q*, which enters into the computation of the overall importance of country *a*, in a way consistent with the definition of weights as importance measures.

In the TAI example, the pair-wise comparison of such as Finland and the United States shows that Finland has better scores for the sub-indicators--*Internet* (weight 1/8), *Telephones* (weight 1/8), *Electricity* (weight 1/8) and *University* (weight 1/8). Thus the score for Finland is 4*1/8=0.5, while the complement to one is the score of the US. The resulting outranking matrix is (**Table 26**):

Table 26. Outranking impact matrix for TAI (five countries)

	Finland	USA	Sweden	Japan	Korea
Finland	0	0.5	0.375	0.75	0.625
USA	0.5	0	0.5	0.625	0.625
Sweden	0.625	0.5	0	0.75	0.625
Japan	0.25	0.375	0.25	0	0.75
Korea	0.375	0.375	0.375	0.25	0

The way the information are combined generates several possible ranking procedures (Young, 1988; Munda, 2004), each with pros and cons. One possible algorithm is the Condorcet-Kemeny-Young-Levenglick (CKYL) ranking procedure (Munda and Nardo 2003). According to CKYL, the ranking of countries with the highest likelihood is the one supported by the maximum number of sub-indicators for each pair-wise comparison, summed over all pairs of countries considered. More formally, all the $M(M-1)$ pair-wise comparisons compose the outranking matrix *E*. Call *R* the set of all *M!* possible complete rankings of alternatives, $R=\{r_S\}$, $s=1,2,..., M!$. For each r_S, compute the corresponding score φ_s as the summation of e_{jk} over all the $\binom{M}{2}$ pairs j,k of alternatives.

That is, $\varphi_s = \sum e_{jk} \quad where\ j \neq k,\ s = 1,2,...M! \quad and \quad e_{jk} \in r_s$ .The final ranking ($r*$) is the solution of:

$$r* \Leftrightarrow \varphi_* = \max \sum e_{jk} \qquad where\ e_{jk} \in R \tag{32}$$

In the TAI example, the number of permutations obtained from 5 countries are 120, the first 5 are listed in **Table 27**. For example, the score of the first ranking (USA, Sweden, Finland, Japan and Korea) is obtained as follows: according to the impact matrix the comparison of US with the other countries yields

0.5 against Finland and Sweden, and 0.625 against Japan and Korea (overall 2.25). The comparison of Sweden yields 0.625 against Finland and Korea and 0.75 against Japan (overall 2). Finland obtains 0.625 against Korea and 0.75 against Japan (overall 1.375). Finally Japan obtains 0.75 against Korea. The final score of this ranking is then equal to 2.25+2+1.375+0.75=6.375.

Table 27. Permutations obtained from the outranking matrix for TAI and associated score

USA	Sweden	Finland	Japan	Korea	**6.375**
Sweden	Finland	USA	Japan	Korea	**6.375**
Sweden	USA	Finland	Japan	Korea	**6.375**
Finland	USA	Sweden	Japan	Korea	6.125
Finland	Sweden	USA	Japan	Korea	6.125
USA	Finland	Sweden	Japan	Korea	6.125

According to expression (32) the final ranking will be the permutation(s) with the highest score. In our example the first 3 permutations have the highest overall score, and thus all those can be considered as a winning ranking.

This aggregation method has the advantage to overcome some of the problems raised by additive or multiplicative aggregations, e.g., preference dependence, the use of different ratio or interval scale to express the same indicator and the meaning of trade-offs given to the weights. With this method, moreover, qualitative and quantitative information can be jointly treated. In addition, it does not need any manipulation or normalisation to assure the comparability of sub-indicators. The drawbacks on the other hand include the dependence of irrelevant alternatives, i.e. the possible presence of cycles/rank reversal in which in the final ranking, country *a* is preferred to *b*, *b* is preferred to *c* but *c* is preferred to *a* (the same problem highlighted for AHP with indicators). Furthermore, information on intensity of preference of variables is never used: if one indicator for country *a* is much less than the same indicator for country *b* produces the same ranking as the case in which this difference is very small[29]. Notice that with this method, the focal point is shifted to the determination of weights, which is crucial for the result[30].

Geometric aggregation

An undesirable feature of additive aggregations is the implied full compensability, such that poor performance in some indicators can be compensated by sufficiently high values of other indicators. For example if an hypothetical composite were formed by inequality, environmental degradation, GDP per capita and unemployment, two countries, one with values 21, 1, 1, 1; and the other with 6,6,6,6 would have equal composite if the aggregation is additive and EW is applied. Obviously the two countries would represent very different social conditions that would not be reflected in the composite. If multi-criteria analysis entails full non-compensability, the use of a geometric aggregation (also called deprivational index) $CI_c = \prod_{q=1}^{Q} x_{q,c}^{w_q}$ is an in-between solution[31].

In the example above, the first country would have a much lower composite than the second, if the aggregation is geometric (2.14 for the first and 6 for the second). In a benchmarking exercise, countries with low scores in some sub-indicators thus would prefer a linear rather than a geometric aggregation. On the other hand, the marginal utility of an increase in the score would be much higher when the absolute value of the score is low: the first country increasing the second indicator by 1 unit would increase its composite from 2.14 to 2.54, while country 2 would go from 6 to 6.23. In other terms, the first country would increase its composite by 19% while the second only by 4%. Consequently, a country should be more interested in increasing those sectors/activities/alternatives with the lowest score in order to have a higher chance to improve its position in the ranking, if the aggregation is geometric rather than linear (Zimmermann and Zysno, 1983).

Furthermore, the type of aggregation employed is strongly related with the method used to normalise raw data (*Step 5*). In particular, Ebert and Welsch (2004) has shown that the use of linear aggregations yields meaningful composite indicators, only if data are all expressed in partially comparable interval scale (i.e. temperature in Celsius of Fahrenheit) of type $f : x \rightarrow \alpha x + \beta_i \quad \alpha > 0$ (i.e. α fixed, but β_i varying across subindicators) or in a fully comparable interval scale (β constant); Non-comparable data measured in ratio scale (i.e. kilograms and pounds) $f : x \rightarrow \alpha_i x$ where $\alpha_i > 0$ (i.e. α_i varying across sub-indicators) can only be meaningfully aggregated by using geometric functions, provided that *x* is strictly positive. In other terms, except in the case of all indicators measured in different ratio scale, the measurement scale must be the same for all indicators when aggregating. Thus, care should be given when indicators measured in different scale coexist in the same composite. The normalisation method should be properly used to remove the scale effect.

Table 28 highlights the dependence of rankings to the aggregation methods used (in this case linear, geometric and based on the multi-criteria technique for the TAI dataset with 23 countries). Although in all cases, equal weighting is used, the resulting rankings are very different. For example, Finland ranks first according to the linear aggregation, second according to the geometric aggregation and third according to the multi-criteria. Note that Korea ranks sixteenth with GME, while its ranking is much higher according to the other two methods, while the reverse is true for Belgium.

Table 28. TAI country rankings by different aggregation methods

	Position in the ranking		
	LIN	**NCMC**	**GME**
Finland	1	3	2
United States	2	1	1
Sweden	3	2	3
Japan	4	4	4
Korea, Rep. of	5	9	16
Netherlands	6	8	5
United Kingdom	7	5	6
Singapore	8	12	18
Canada	9	11	13
Australia	10	9	14
Germany	11	7	8
Norway	12	6	11
Ireland	13	13	7
Belgium	14	17	9
New Zealand	15	15	17
Austria	16	15	12
France	17	14	10
Israel	18	18	15
Spain	19	20	19
Italy	20	19	21
Czech Republic	21	21	23
Hungary	22	23	22
Slovenia	23	22	20

Note: Dataset TAI, for 23 countries. Numbers refer to the position in ranking.

UNCERTAINTY AND SENSITIVITY ANALYSIS

Sensitivity analysis for modelers? Would you go to an orthopaedist who didn't use X-rays? This sentence by J.M. Furbringer is the essence of what sensitivity analysis is and why it is useful. The composite indicators development involves stages where subjective judgment has to be made: the selection of sub-indicators, the treatment of missing values, the choice of aggregation model, the weights of the indicators, etc. All these subjective choices are the *bones* of the composite indicator and shape, together with the information provided by the numbers themselves, the message communicated by the composite indicator. However, composite indicators can send misleading or non-robust policy messages if they are poorly constructed or misinterpreted.

Since the quality of a model also depends on the soundness of its assumptions, good modeling practice requires that the modeler provides an evaluation of the confidence in the model, assessing the uncertainties associated to the modeling process and the subjective choices undertaken. This is what sensitivity analysis does: it performs the 'X-rays' of the model by studying the relationship between information flowing in and out of the model.

More formally, sensitivity analysis is the study of how the variation in the output can be apportioned, qualitatively or quantitatively, to different sources of variation in the assumptions, and of how the given composite indicator depends upon the information fed into it. Sensitivity analysis is thus closely related to uncertainty analysis which aims to quantify the overall uncertainty in the countries' ranking as a result of the uncertainties in the model input. A combination of uncertainty and sensitivity analysis can help to gauge the robustness of the composite indicator ranking, to increase its transparency, to identify which countries are favoured or deteriorate under certain assumptions and to help framing a debate around the Index.

Next we describe how to apply uncertainty and sensitivity analysis to composite indicators. Our synergistic use of uncertainty and sensitivity analysis has recently been applied for the robustness assessment of composite indicators (Saisana *et al.* 2005a, Saltelli *et al.*, 2004) and has proven to be useful in dissipating some of the controversy surrounding composite indicators such as the Environmental Sustainability Index (Saisana *et al.* 2005b).

In the TAI case study we focus on 5 main uncertainties/assumptions: inclusion-exclusion of one indicator at-a-time, imputation of missing data, different normalisation methods, different weighting schemes and different aggregation schemes.

Let *CI* be the index value for country *c, c=1,…,M,*

$$CI_c = f_{rs}\left(I_{1,c}, I_{2,c}, \dots I_{Q,c}, w_{s,1}, w_{s,2}, \dots w_{s,Q}\right) \qquad (33)$$

according the weighting model f_{rs}, $r = 1,2,3, \quad s = 1,2,3$, where the index *r* refers to the aggregation system (LIN, GME, NCMC) and index *s* refers to the weighting scheme (BAL, AHP, BOD). The index is based on *Q* normalised sub-indicators $I_{1,c}, I_{2,c}, \dots I_{Q,c}$ for that country and scheme-dependent weights $w_{s,1}, w_{s,2}, \dots w_{s,Q}$ for the sub-indicators. The most frequently used normalisation methods for the sub-indicators are based on the re-scaled (34a), standardised (34b), or on the raw indicator values (34c).

$$\begin{cases} I_{q,c} = \dfrac{x_{q,c} - \min(x_q)}{range(x_q)} & \text{(34a)} \\ I_{q,c} = \dfrac{x_{q,c} - mean(x_q)}{std(x_q)} & \text{(34b)} \\ I_{q,c} = x_{q,c} & \text{(34c)} \end{cases}$$

where $I_{q,c}$ is the normalised and $x_{q,c}$ is the raw value of the sub-indicator xq for country c.

Note that the re-scaled value (34a) can be used in conjunction with all the weighting schemes (BAL, AHP and BOD) for all aggregation systems (LIN, GME, NCMC). The standardised value (34b) can be used with weighting schemes (BAL, AHP) for aggregation systems (LIN, NCMC). And the raw indicator value (34c) can be used with weighting schemes (BAL, AHP) for aggregation systems (GME, NCMC).

The rank assigned by the composite indicator to a given country, i.e. $Rank(CI_c)$ is an output of the uncertainty/sensitivity analysis. The average shift in country rankings is also explored. This latter statistic captures the relative shift in the position of the entire system of countries in a single number. It can be calculated, as the average of the absolute differences in countries' rank with respect to a reference ranking over the *M* countries:

$$\overline{R}_S = \frac{1}{M}\sum_{c=1}^{M}\left|Rank_{ref}(CI_c) - Rank(CI_c)\right| \qquad (35)$$

The reference ranking for the TAI analysis is the original rank given to the country by the original version of the index. The investigation of $Rank(CI_c)$ and $\overline{R}_S$ is the scope of the uncertainty and sensitivity analysis[32].

General framework

The analysis is conducted as a single Monte Carlo experiment, e.g. by eliminating all uncertainty sources simultaneously to capture all possible synergy effects among uncertain input factors. This involves the use of triggers, e.g. the use of uncertain input factors to decide which aggregation system and weighting scheme to adopt. A discrete uncertain factor, which can take integer values between 1 and 3 is used for the aggregation system and similarly for the weighting scheme. Other trigger factors are generated to select indicators to be omitted, the editing scheme, the normalisation scheme and so on, until a full set of input variables is available to compute $Rank(CI_c)$, $\overline{R}_S$.

Uncertainty analysis (UA)

Various components of the CI construction process can introduce uncertainty in the output variables, $Rank(CI_c)$ and $\overline{R}_S$. The UA is essentially based on simulations that are carried on various equations that constitute the underlying *model*. The uncertainties are transferred into a set of scalar input factors, such that the resulting $Rank(CI_c)$ and $\overline{R}_S$ are non-linear functions of the uncertain input factors, and the estimated probability distribution (pdf) of $Rank(CI_c)$ and $\overline{R}_S$. Various methods are available for evaluating output uncertainty. The following is the Monte Carlo approach, which is based on multiple evaluations of the model with k randomly selected model input factors. The procedure has six steps:

Step 1. Assign a pdf to each input factor X_i, $i = 1,2...k$. The first input factor, X_1 is used for the selection of the editing scheme (for the second TAI analysis only):

X_1	Estimation of missing data
1	Use bivariate correlation to impute missing data
2	Assign zero to missing datum

The second input factor X_2 is the trigger to select the normalisation method.

X_2	Normalisation
1	Rescaling (Equation 7.3a)
2	Standardisation (Equation 7.3b)
3	None (Equation 7.3c)

Both X_1 and X_2 are discrete random variables. In practice, they are generated by drawing a random number ζ, uniformly distributed between [0,1] and applying the so called *Russian roulette algorithm*, e.g. for X_1, select 1 if $\zeta \in [0,0.5)$ and 2 if $\zeta \in [0.5,1]$. Uncertain factor X_3 is generated to select which sub-indicator, if any, should be omitted.

ζ	X_3, excluded sub-indicator
$[0, \frac{1}{Q+1})$	None ($X_3 = 0$) all subindicators are used)
$[\frac{1}{Q+1}, \frac{2}{Q+1})$	$X_3 = 1$
…	…
$[\frac{Q}{Q+1}, 1]$	$X_3 = Q$

That is with probability $\frac{1}{Q+1}$ no sub-indicator will be excluded, while with probability $[1-\frac{1}{Q+1}]$ one of the Q sub-indicators will be excluded with equal probability. Clearly, one could have made the probability of $X_3 = 0$ larger or smaller than $\frac{1}{Q+1}$ and still sample the values $X_3 = 1,2,...Q$ with equal probability. A scatter-plot based sensitivity analysis would be used to track which indicator affects the output the most when excluded. Also recall that whenever a sub-indicator is excluded, the weights of the other factors are re-scaled to 1 to make the composite index comparable if either BAL or AHP. When BOD is selected the exclusion of sub-indicator leads to a re-execution of the optimisation algorithm.

Trigger X_4 is used to select the aggregation system

X_4	Aggregation Scheme
1	LIN
2	GME
3	NCMC

Note that when LIN is selected the composite indicators are computed as:

$$CI_c = \sum_{q=1}^{Q} w_{sq} I_{q,c} \tag{36}$$

while when GME is selected they are:

$$CI_c = \prod_{q=1}^{Q} \left(I_{q,c}\right)^{w_{4,q}} \tag{37}$$

When NCMC is selected the countries are ranked directly from the outscoring matrix.

X_5 is the trigger to select the weighting scheme;

X_5	Weighting Scheme
1	BAL
2	AHP
3	BOD

The last uncertain factor X_6 is used to select the expert. In this experiment, there are 20 experts. Once an expert is selected at runtime via the trigger X_6, the weights assigned by that expert (either for the BAL or AHP schemes) are assigned to the data. Clearly the selection of the expert has no bearing when BOD is selected ($X_5 = 3$). All the same this uncertain factor would be generated at each individual Monte Carlo simulation, given the row dimension of the Monte Carlo sample (*constructive dimension*) should be fixed in a Monte Carlo experiment, i.e. even if some of the sampled factors will not be active at a particular run, they will be all the same generated by the random sample generation algorithm. The constructive dimension of the Monte Carlo experiment, the number of random numbers to be generated for each trial, is hence $k = 6$ .Note that alternative arrangements of the analysis would have been possible.

Step 2. Generate randomly N combinations of independent input factors $\mathbf{X}^l$, $l = 1,2,...N$ (a set $\mathbf{X}^l = X_1^l, X_2^l, \ldots, X_k^l$ of input factors is called a sample). For each trial sample $\mathbf{X}^l$ the computational model can be evaluated, generating values for the scalar output variable Y^l, where Y^l is either $Rank(CI_c)$, the value of the rank assigned by the composite indicator to each country, or $\overline{R}_S$, the averaged shift in countries' rank.

Step 3. Close the loop over *l*, and analyse the resulting output vector $\boldsymbol{Y}^l$, with $l = 1, ..., N$.

The generation of samples can be performed using various procedures, such as simple random sampling, stratified sampling, quasi-random sampling or others (Saltelli *et. al*, 2000a). The sequence of $\boldsymbol{Y}^l$ gives the pdf of the output *Y*. The characteristics of this pdf, such as the variance and higher order moments, can be estimated with an arbitrary level of precision that is related to the size of the simulation N .

Sensitivity analysis using variance-based techniques

A necessary step when designing a sensitivity analysis is to identify the output variables of interest. Ideally these should be relevant to the issue tackled by the model (Saltelli *et al.*, 2000b, 2004). It has been noted earlier that composite indicators can be considered as models. When several layers of uncertainty are simultaneously present, composite indicators could become a non-linear, possibly non-additive model. As argued by practitioners (Saltelli et al., 2000a, EPA, 2004), for non-linear models, robust, "model-free" techniques should be used for sensitivity analysis. Sensitivity analysis using variance-based techniques are model free and display additional properties convenient for the present analysis, such as:

- they allow an exploration of the whole range of variation of the input factors, instead of just sampling factors over a limited number of values, e.g. in fractional factorial design (Box et al. 1978);
- they are quantitative, and can distinguish main effects (first order) from interaction effects (higher order);
- they are easy to interpret and to explain;
- they allow for a sensitivity analysis whereby uncertain input factors are treated in groups instead of individually;
- they can be justified in terms of rigorous settings for sensitivity analysis.

To compute a variance based sensitivity measure for a given input factor X_i, start from the fractional contribution to the model output variance, i.e. the variance of Y where Y is either $Rank(CI_c)$, and $\overline{R}_S$) due to the uncertainty in X_i :

$$V_i = V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i)) \tag{38}$$

Fix factor X_i, e.g. to a specific value x_i^* in its range, and compute the mean of the output Y averaging over all factors but factor X_i: $E_{\mathbf{X}_{-i}}(Y|X_i = x_i^*)$. Then, take the variance of the resulting function of x_i^* over all possible x_i^* values. The result is given by Equation (38), where the dependence from x_i^* has been dropped. V_i is a number between 0 (when X_i does not gives a contribution to Y at the first order), and $V(Y)$, the unconditional variance of Y, when all factors other than X_i are non influential at any order. Note that the following is always true:

$$V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i)) + E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i)) = V(Y) \tag{39}$$

where the first term (39) is called a main effect, and the second one the residual. An important factor should have a small residual, e.g. a small value of $E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i))$. This is intuitive as the residual measures the expected reduced variance that one would achieve if one could fix X_i. Rewrite this as $V_{\mathbf{X}_{-i}}(Y|X_i = x_i^*)$, a variance conditional on x_i^*. Then the residual $E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i))$ is the expected value of such conditional variance, averaged over all possible values of x_i^*. This would be small if X_i is influential. A first order sensitivity index is obtained through normalising the first-order term by the unconditional variance:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))}{V(Y)} = \frac{V_i}{V(Y)} \tag{40}$$

One can compute conditional variances corresponding to more than one factor, e.g. for two factors X_i and X_j, the conditional variance would be $V_{X_iX_j}(E_{\mathbf{X}_{-ij}}(Y|X_i, X_j))$, and the second-order term variance contribution would become:

$$V_{ij} = V_{X_iX_j}(E_{\mathbf{X}_{-ij}}(Y|X_i, X_j)) - V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i)) - V_{X_j}(E_{\mathbf{X}_{-j}}(Y|X_j)) \tag{41}$$

where clearly V_{ij} is different from zero only if $V_{X_iX_j}(E_{\mathbf{X}_{-ij}}(Y|X_i, X_j))$ is larger than the sum of the first-order term relative to factors X_i and X_j.

When all *k* factors are independent from one another, the sensitivity indices can be computed using the following decomposition formula for the total output variance *V(Y)* :

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \sum_i \sum_{j>i} \sum_{\substack{l>j \\ j>i}} V_{ijl} + \ldots + V_{12\ldots k} \tag{42}$$

Terms above the first order (42) are known as interactions. A model without interactions among its input factors is said to be *additive*. In this case, $\sum_{i=1}^{k} V_i = V(Y)$, $\sum_{i=1}^{k} S_i = 1$ and the first order conditional variances of equation (38) are all needed to decompose the model output variance. For a non-additive model, higher order sensitivity indices, responsible for interaction effects among sets of input factors, have to be computed. However, higher order sensitivity indices are usually not estimated, as in a model with *k* factors, the total number of indices (including the S_i 's) that needs to be estimated would be as high as 2^k-1. Instead a more compact sensitivity measure is used. The total effect sensitivity index concentrates on a single term for all the interactions, involving a given factor X_i. To give an example, for a model of *k*=3 independent factors, the three total sensitivity indices would be:

$$S_{T1} = \frac{V(Y) - V_{X_2X_3}(E_{X_1}(Y|X_2, X_3))}{V(Y)} = S_1 + S_{12} + S_{13} + S_{123} \tag{43}$$

And analogously:

$$S_{T2}=S_2+S_{12}+S_{23}+S_{123} \tag{44}$$
$$S_{T3}=S_3+S_{13}+S_{23}+S_{123}$$

The conditional variance $V_{X_2X_3}(E_{X_1}(Y|X_2,X_3))$ in equation (43) can be written in general terms as $V_{\mathbf{X}_{-i}}(E_{X_i}(Y|\mathbf{X}_{-i}))$ (Homma and Saltelli, 1996). This is the total contribution to the variance of *Y* due to non-X_i, i.e., to the *k*-1 remaining factors, such that $V(Y)-V_{\mathbf{X}_{-i}}(E_{X_i}(Y|\mathbf{X}_{-i}))$ includes all terms. In general, $\sum_{i=1}^{k} S_{Ti} \geq 1$.

Given (39), the total effect sensitivity index can also be written as:

$$S_{Ti} = \frac{V(Y)-V_{\mathbf{X}_{-i}}(E_{X_i}(Y|\mathbf{X}_{-i}))}{V(Y)} = \frac{E_{\mathbf{X}_{-i}}(V_{X_i}(Y|\mathbf{X}_{-i}))}{V(Y)} \tag{45}$$

For a given factor, X_i ,a significant difference between S_{Ti} and S_i signals an important role of interaction for that factor in *Y*. Highlighting interactions among input factors helps to improve our understanding of the model structure. Estimators for both (S_i, S_{Ti}) are provided by a variety of methods reviewed in Chan *et al.* (2000). Here the method of Sobol' (1993), in its improved version due to Saltelli (2002) is used. The method of Sobol' uses quasi-random sampling of the input factors. The pair (S_i, S_{Ti}) gives a fairly good description of the model sensitivities, which for the improved Sobol' method is of 2*n*(*k*+1) model evaluations, where *n* represents the sample size, required to approximate the multidimensional integration implicit in the E and V operators above to a plain sum. *n* can vary in the hundred-to-thousand range.

When the uncertain input factors X_i are dependent, the output variance cannot be decomposed, as in equation (42). The S_i, S_{Ti} indices, defined by (38) and (45) are still valid sensitivity measures for X_i, though their interpretation has changed, e.g. S_i could carry over the effects of other factors that can be positively or negatively correlated to X_i (see Saltelli and Tarantola, 2002), while S_{Ti} can no longer be decomposed meaningfully into main effect and interaction effects. The S_i, S_{Ti}, for the case of non-independent input factors, could also be interpreted in as "settings" for sensitivity analysis.

A description of two settings linked to S_i, S_{Ti} is discussed below[33].

Factors' Prioritisation (FP) Setting. Suppose a factor that, once "discovered" in its true value and fixed, would reduce the most V(Y). The true values for the factors however are unknown. The best choice one can make would be the factor with the highest S_i, whether the model is additive, and the factors are independent or not.

Factors' Fixing (FF) Setting. Can one fix a factor [or a subset of input factors] at any given value over their range of uncertainty without reducing significantly the variance of the output? One can only fix those (sets of) factors whose S_{Ti} is zero.

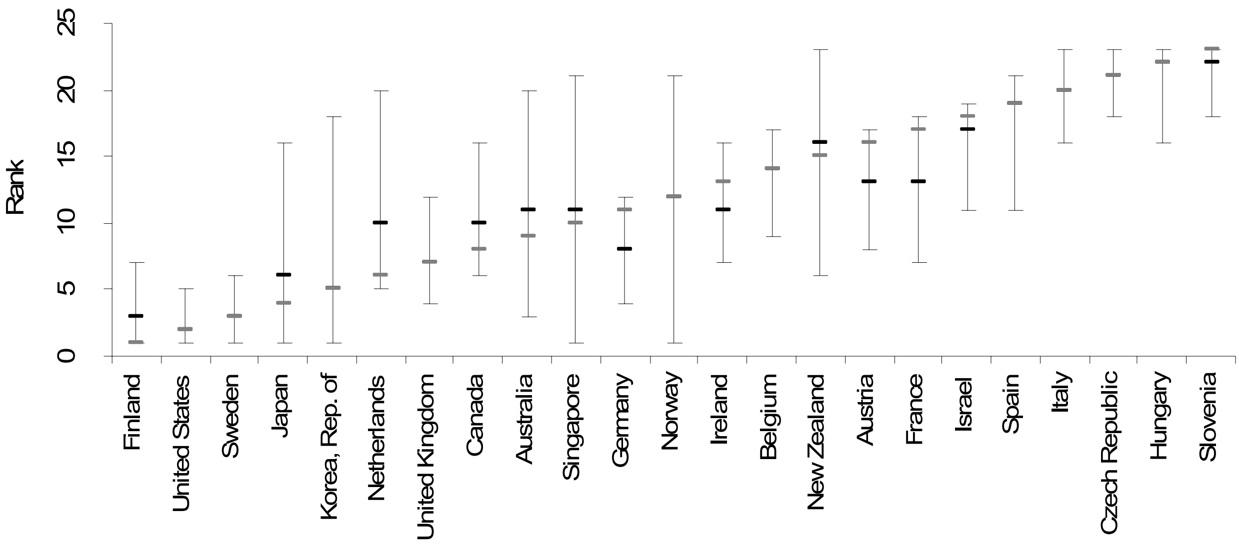
The extended variance-based methods, including the improved version of Sobol', for both dependent and independent input factors, are implemented in the freely distributed software SIMLAB (Saltelli *et al.*, 2004).

Analysis 1

The first analysis is run without imputation, i.e. by censoring all countries with missing data. As a result, only 34 countries in theory could be analysed. Other countries from rank (original TAI) 24 are also dropped, e.g., Hong Kong, as this is the first country with missing data. The analysis is restricted to the set of countries whose rank is not altered by the omission of missing records. The uncertainty analysis for the remaining 23 countries is given in **Figure 18** for the ranks, with countries ordered by their original TAI position, ranging from Finland, rank=1, to Slovenia, rank=23. Note that the choice of ranks, instead of composite indicator values, is dictated by the use of the NCMC aggregation system.

The width of the 5th – 95th percentile bounds and the ordering of the medians (black hyphen) often are at odds with the ordering of the original TAI (grey hyphen). Although one could still see the difference between the group of leader and that of laggards, there are considerable differences between the new and the original TAI. If the uncertainty within the system were a true reflection of the status of knowledge and the (lack of) consensus among experts on how TAI should be built, we would have to conclude that TAI is not a robust measure of country technology achievement.

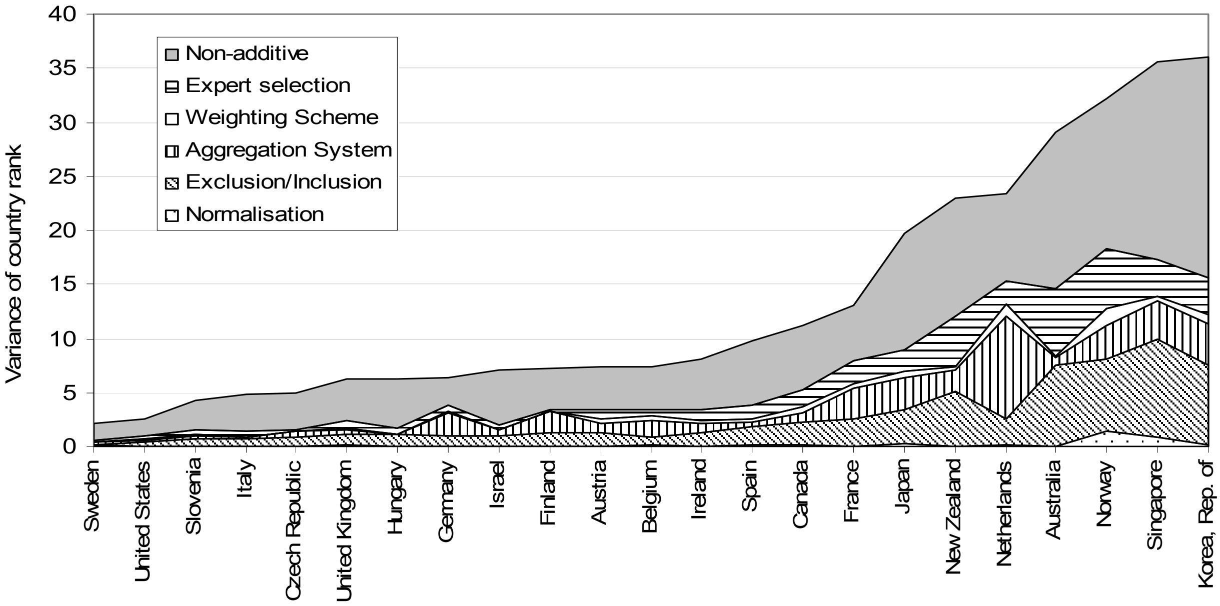
Figure 18. Uncertainty analysis of TAI country rankings



Note: Results show the country rankings according to the original TAI 2001 (light grey marks), and the median (black mark) and the corresponding 5th and 95th percentiles (bounds) of the distribution of the MC-TAI for 23 countries. Uncertain input factors: normalisation method, inclusion-exclusion of a sub-indicator, aggregation system, weighting scheme, expert selection. Countries are ordered according to the original TAI values.

Figure 19 shows the sensitivity analysis based on the first order indices calculated by the Sobol' (1993) method and the improved version due to Saltelli (2002). The total variance for each country's rank is presented along with the part that can be decomposed according to the first order conditional variances. The aggregation system, followed by the inclusion-exclusion of sub-indicators and expert selection are the most influential input factors. The countries with the highest total variance in ranks are the middle-performing countries, while the leaders and laggards in technology achievement have low total variance. The non-additive, non-linear part of the variance that is not explained by the first order sensitivity indices ranges from 35% for the Netherlands to 73% for the United Kingdom, whilst for most countries it exceeds 50%. This underlines the necessity of computing higher order sensitivity indices that capture the interaction effect among the input factors.

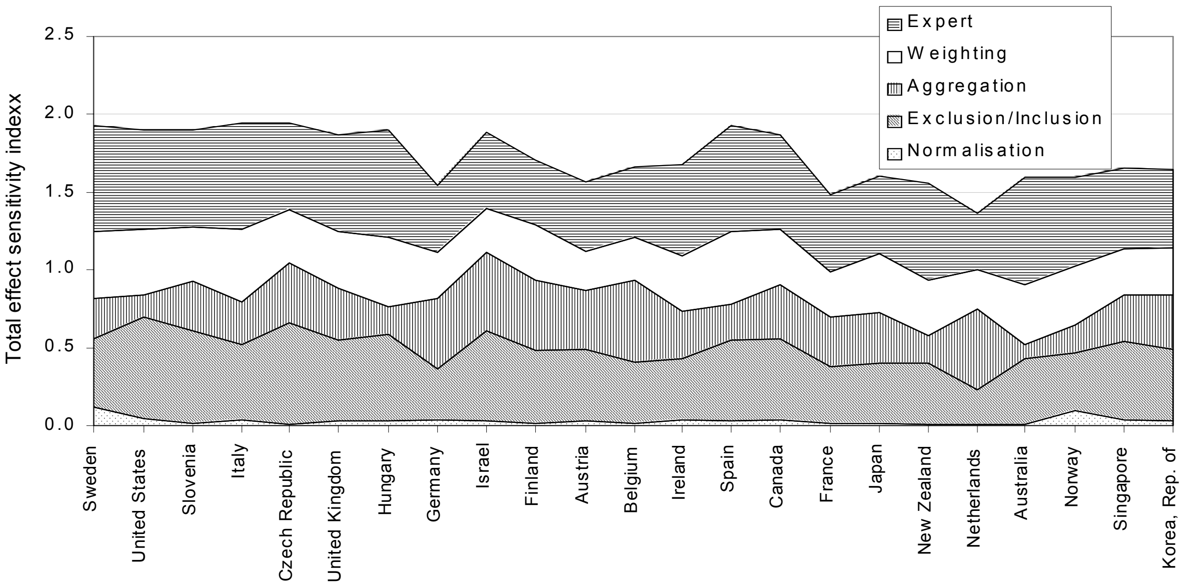
Figure 19. Sobol' sensitivity measures of first orderTAI results



Note: Results are based on the first order indices. Decomposition of country's variance according to the first order conditional variances. Aggregation system, followed by the inclusion-exclusion of sub-indicator and expert selection are the most influential input factors. The part of the variance that is not explained by the first order indices is noted as non-additive. Countries are ordered in ascending order of total variance.

Figure 20 shows the total effect sensitivity indices for the variance of each country's rank. The total effect sensitivity indices concentrate on one single term for all the interactions involving each input factor. The indices add up to a number greater than one due to existing interactions, which seem to exist among the identified influential factors.

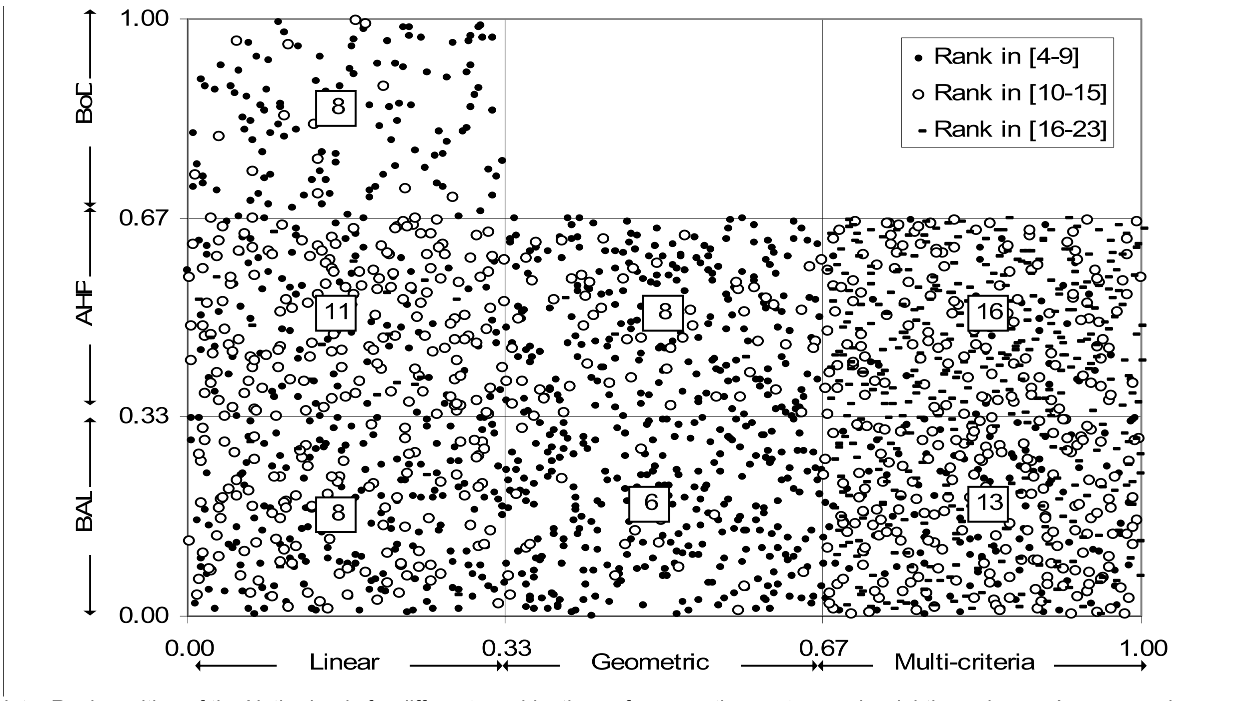
Figure 20. Sobol' sensitivity measures of TAI total effect indices



Note: Results are based on the total effect indices. Aggregation system inclusion-exclusion of sub-indicator and expert selection present most of the interaction effects. Countries are ordered in ascending order of total variance.

If the TAI model were additive with no interactions between the input factors, the non-additive part of the variance in Figure 21 would have been zero. In other words, the first order sensitivity indices would have summed to 1, and the sum of the total effect sensitivity indices in Figure 22 would have been 1. Yet, the sensitivity indices show the high degree of non-linearity and additivity for the TAI model and the importance of the interactions. The high effect of interactions for the Netherlands, which also has a large percentile bound, is further explored. **Figure 21** shows that the Netherlands is favoured by the combination of “geometric mean system” with “BAL weighing”, and not favoured by the combination of “Multi criteria system” with “AHP weighting”. This is a clear interaction effect. In depth analysis of the output data reveals that as far as inclusion – exclusion is concerned, it is the exclusion of the sub-indicator “Royalties” leading to worsening of the Netherlands' ranking under any aggregation system.

Figure 21. Netherlands ranking by aggregation and weighting systems



Note: Rank position of the Netherlands for different combinations of aggregation system and weighting scheme. Average rank per case is indicated in the box. The interaction effect between aggregation system and weighting scheme is clear.

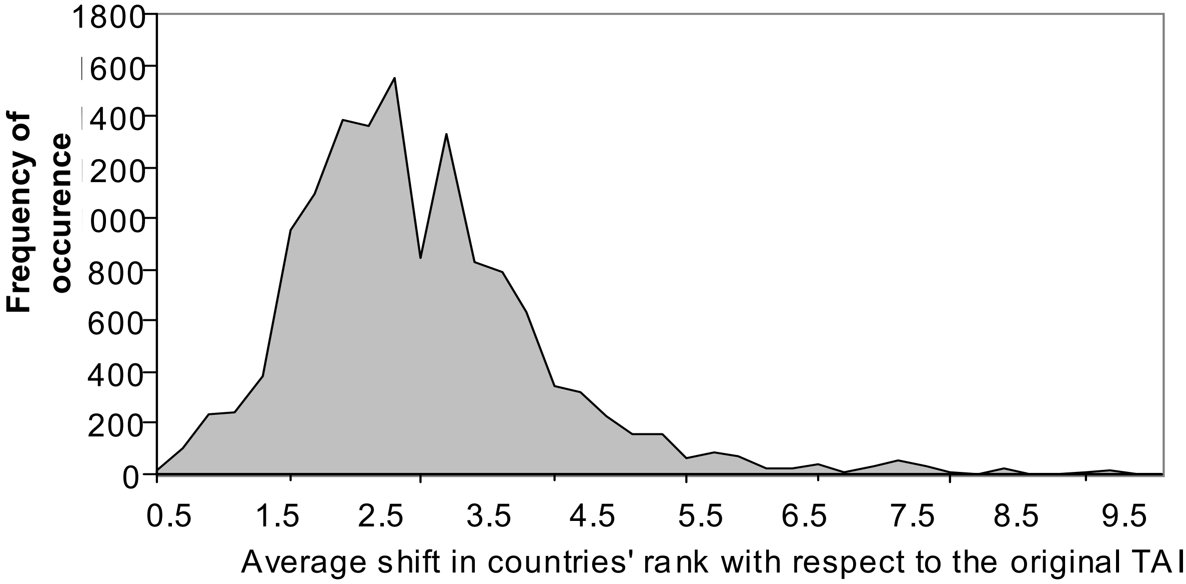
Figure 22 shows the histogram of values for the average shift in the rank of the output variable (Equation 7.2) with respect to the original TAI rank. The mean value is almost 3 positions, with a standard deviation slightly above 1 position. The input factors -- the aggregation system plus inclusion/exclusion at the first order -- affect this variable the most (**Table 29**). When the interactions are considered, both weighting scheme and expert choice become important. This effect can be seen in **Figure 23**. In some cases the average shift in country’s rank when using NCMC can be as high as 9 places.

Table 29. Sobol’ sensitivity measures of first order and total effects on TAI results

Input Factors	*First order (S_i)*	*Total effect (S_{Ti})*	S_{Ti} - S_i
Normalisation	0.000	0.008	0.008
Exclusion/Inclusion of sub-indicator	0.148	<u>0.435</u>	<u>0.286</u>
Aggregation system	<u>0.245</u>	0.425	0.180
Weighting Scheme	0.038	<u>0.327</u>	<u>0.288</u>
Expert selection	0.068	<u>0.402</u>	<u>0.334</u>
Sum	0.499	1.597	

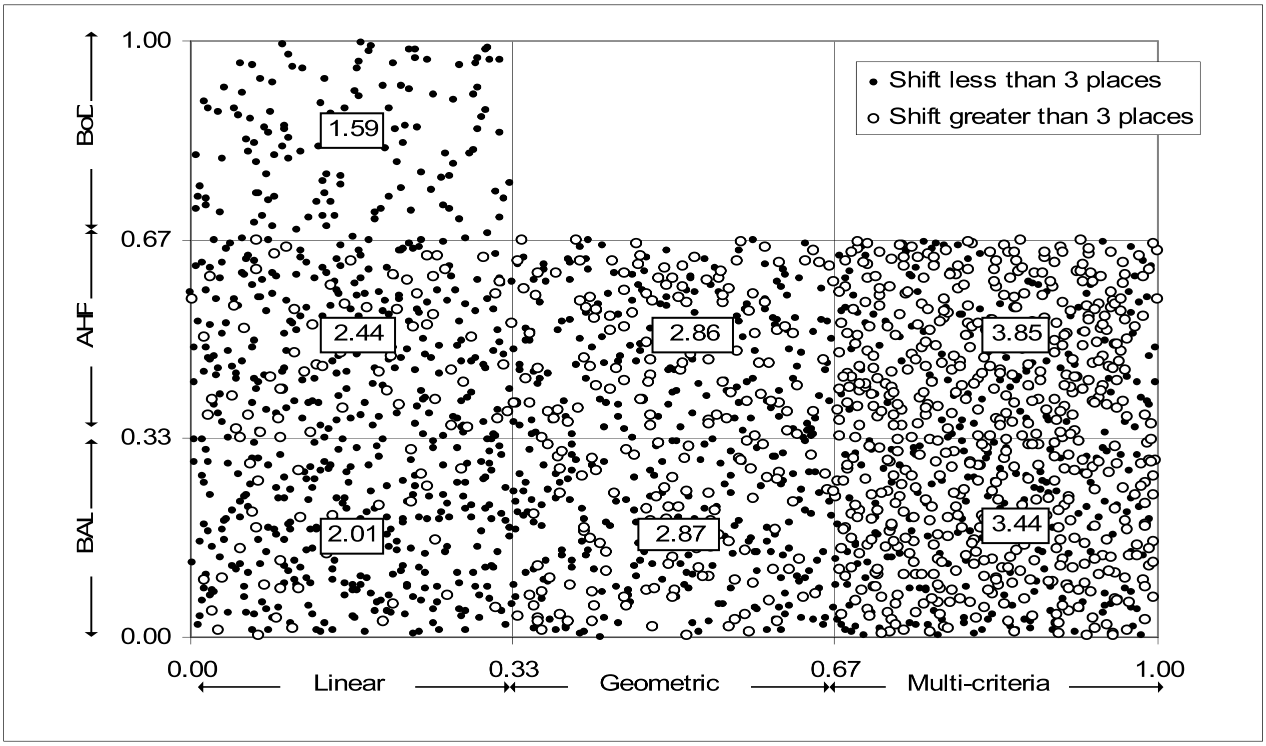
Note: Average shift in countries’ rank with respect to the original TAI . Significant values are underlined.

Figure 22. Uncertainty analysis for TAI output variable



Note: Average shift in countries' rank with respect to the original TAI. Uncertain input factors: normalisation method, inclusion-exclusion of a sub-indicator, aggregation system, weighting scheme, expert selection.

Figure 23. Average shift in TAI country rankings by aggregation and weighting combinations



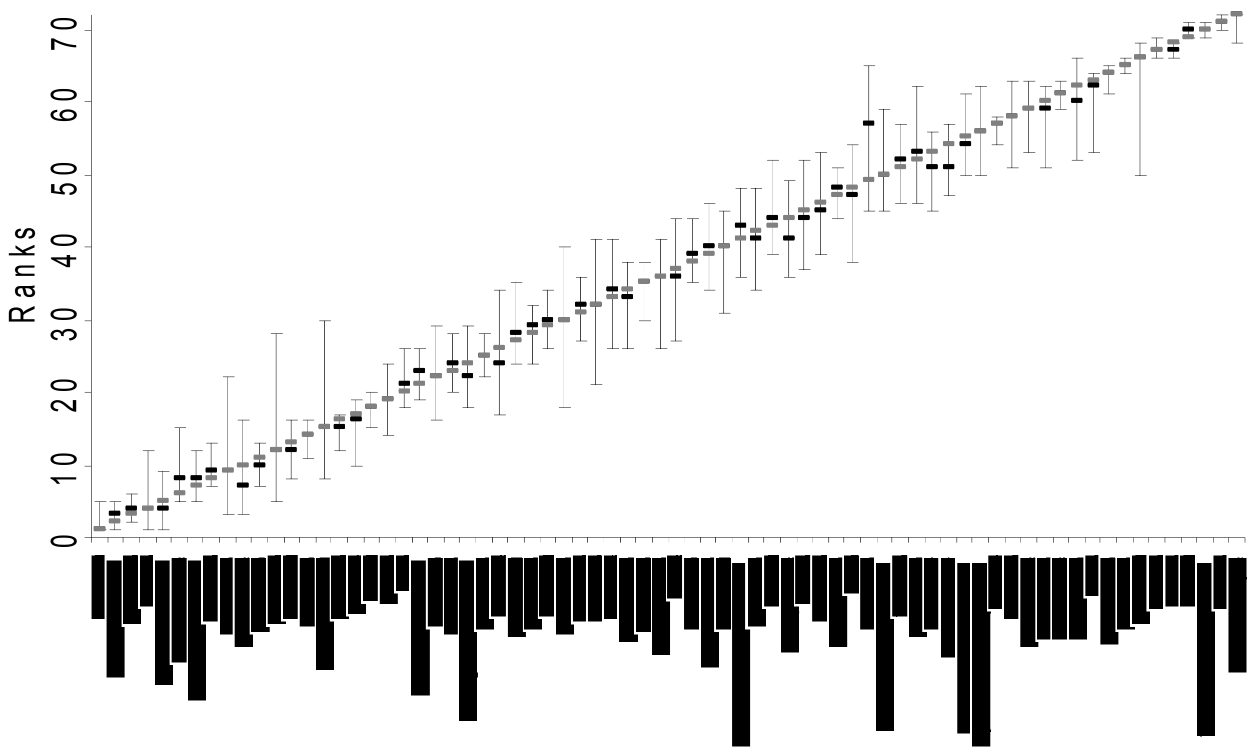
Note: Average shift in countries' rank with respect to the original TAI for different combinations of aggregation system and weighting scheme. Average value per case is indicated in the box.

Analysis 2

In this analysis, it is assumed that the TAI stakeholders have agreed on a linear aggregation system. In fact, one might argue that the choice of the aggregation system is to some extent dictated by the use of the index, and by the expectation of its stakeholders. For instance, if stakeholders believe that the system should be non-compensatory, NCMC would be adopted. Eventually, this would lead to an on average medium–good performance, being worth more for a country than a performance, which is very good on some sub-indicators and bad in others. A GME approach would follow the progress of the index overtime in a scale-independent fashion.

Given these considerations, the second analysis is based on the LIN system, as in the original TAI. The uncertainty analysis plot (**Figure 24**) shows a much more robust behaviour of the index, with fewer inversions of rankings, when median-TAI and original TAI are compared. As far as for the sensitivity, the uncertainty arising from imputation does not seem to make a significant contribution to the output uncertainties, which are also dominated by weighing, inclusion-exclusion, expert selection. Even when, as in the case of Malaysia, imputation by bivariate approach ends into an unrealistic number of patents being imputed for this country (234 patents granted to residents per million people), its rank's uncertainty is still insensitive to imputation. The sensitivity analysis results for the average shift in ranking output variable (Equation 7.2) is shown in **Table 30**. Interactions are now between expert selection and weighing, and considerably less with interaction with inclusion-exclusion.

Figure 24. Uncertainty analysis of TAI country rankings



Note: Uncertainty analysis results showing the countries' rank according to the original TAI 2001 (light grey marks), and the median (black mark) and the corresponding 5th and 95th percentiles (bounds) of the distribution of the MC-TAI for 72 countries. Uncertain input factors: imputation, normalisation method, inclusion-exclusion of a sub-indicator, weighting scheme, expert selection. A linear aggregation system is used. Countries are ordered according to the original TAI values.

Table 30. Sobol' sensitivity measures and average shift in TAI rankings

Input Factors	*First order (S_i)*	*Total effect (S_{Ti})*	S_{Ti} - S_i
Imputation	0.001	0.005	0.004
Normalisation	0.000	0.021	0.021
Exclusion/Inclusion of sub-indicator	0.135	0.214	0.078
Weighting Scheme	<u>0.212</u>	<u>0.623</u>	<u>0.410</u>
Expert selection	<u>0.202</u>	<u>0.592</u>	<u>0.390</u>
Sum	0.550	1.453	

Note: Significant values are underlined.

The use of one strategy versus another in indicator building might lead to a biased picture of the country performance, depending on the severity of the uncertainties. As shown by the preceding analyses, if the constructors of the index disagree on the aggregation system, it is highly unlikely for a robust index to emerge. If uncertainties exist in the context of a well-established theoretical framework, e.g. a participatory approach within a linear aggregation scheme is favoured, the resulting country rankings could be fairly robust in spite of the uncertainties.

Both imputation and normalisation do not affect significantly countries ranking when uncertainties of higher order are present. In the current set-up, the uncertainties of higher order are expert selection and weighing scheme (second analysis). *A fortiori* normalisation does not affect output, when the very aggregation system is uncertain (first analysis). In other words, when the weights are uncertain, it is unlikely that normalisation and editing will affect the country ranks.

The aggregation system is of paramount importance. It is recommended that indicator constructors agree on a common approach. Once the system is fixed, it is the choice of the aggregation methods and of the experts that – together with indicator inclusion – exclusion, dominates the uncertainty in the country ranks. However, note that even in the second analysis, when the aggregation system is fixed, the composite indicator model is strongly non additive, which reinforces the case for the use of quantitative, Monte Carlo based approach to robustness analysis.

REFERENCES

Adriaanse A. (1993), Environmental policy performance. A study on the development of indicators for environmental policy in the Netherlands. SDV Publishers, The Hague.

Anderberg, M.R. (1973), Cluster Analysis for Applications, New York: Academic Press, Inc.

Arrow K.J. (1963), Social choice and individual values, 2d edition, Wiley, New York.

Arrow K.J., and Raynaud H. (1986), Social choice and multicriterion decision making, M.I.T. Press, Cambridge.

Arundel A. and Bordoy C. (2002, Methodological evaluation of DG Research's composite indicators for the knowledge based economy. Document presented by DG RTD at the Inter-service consultation meeting on Structural Indicators on July 11th 2002.

Binder, D.A. (1978), "Bayesian Cluster Analysis," Biometrika, 65, 31 -38.

Boscarino J.A., Figley C.R., and Adams R.E. (2004), "Compassion Fatigue following the September 11 Terrorist Attacks: A Study of Secondary Trauma among New York City Social Workers", International Journal of Emergency Mental Health, Vol. 6, No. 2 • 2004, 1-10.

Box, G., Hunter, W. and Hunter, J. (1978), Statistics for experimenters, New York: John Wiley and Sons.

Bryant F.B., and Yarnold P.R. (1995), Principal components analysis and exploratory and confirmatory factor analysis. In Grimm and Yarnold, Reading and understanding multivariate analysis. American Psychological Association Books.

Chan, K., Tarantola, S., Saltelli, A. and Sobol', I. M. (2000), Variance based methods. In Sensitivity Analysis (eds A. Saltelli, K. Chan, M. Scott) pp. 167-197. New York: John Wiley & Sons.

Charnes A., Cooper W.W., Lewin A.Y., and Seiford L.M. (1995), Data Envelopment Analysis: Theory, Methodology and Applications. Boston:Kluwer.

Cherchye L. (2001), "Using data envelopment analysis to assess macroeconomic policy performance", Applied Economics, 33, 407-416

Cherchye L., and Kuosmanen T. (2002), "Benchmarking sustainable development: a synthetic meta-index approach", EconWPA Working Papers.

Cherchye, L., Moesen W.and Van Puyenbroeck T. (2004), "Legitimately Diverse, Yet Comparable: on Synthesising Social Inclusion Performance in the EU", Journal of Common Market Studies, 42, 919-955.

Commission of the European Communities (1984), The regions of Europe: Second periodic report on the social and economic situation of the regions of the Community, together with a statement of the regional policy committee, OPOCE, Luxembourg.

Cortina, J.M. (1993), " What is coefficient alpha? An examination of theory and applications", Journal of Applied Psychology, 78, 1, 98-104

Cox, D., Fitzpatrick, R., Fletcher, A., Gore, S., Spiegelhalter, D. and Jones, D. (1992), "Quality-of-life assessment: can we keep it simple?", J.R. Statist. Soc. 155 (3), 353-393.

Cronbach, L. J. (1951),"Coefficient alpha and the internal structure of tests.", Psychometrika, 16, 297-334.

Davis, J. (1986); Statistics and Data Analysis in Geology , John Wiley & Sons, Toronto, 646p.

Pan American Health Organization (1996), Annual report of the Director. Healthy People, Healthy Spaces 1996, Official Document No. 283, Washington, D.C. 20037, U.S.A. http://165.158.1.110/english/sha/ops96arx.htm

Debreu G. (1960), Topological methods in cardinal utility theory, in Arrow K.J., Karlin S. and Suppes P. (eds.) Mathematical methods in social sciences, Stanford University Press, Stanford.

Dempster A.P. and Rubin D.B. (1983), Introduction pp.3-10, in Incomplete Data in Sample Surveys (vol. 2): Theory and Bibliography (W.G. Madow, I. Olkin and D.B. Rubin eds.) New York: Academic Press.

Dietz F.J.and van der Straaten J. (1992), "Rethinking environmental economics: missing links between economic theory and environmental policy", Journal of Economic Issues, Vol. XXVI No. 1, pp. 27-51.

Dunteman, G.H. (1989), Principal components analysis, Thousand Oaks, CA: Sage Publications, Quantitative Applications in the Social Sciences Series, No. 69.

European Commission , DG ENTR (2001), European Innovation Scoreboard, Brussels.

Ebert U. and Welsch H. (2004), "Meaningful environmental indices: a social choice approach", Journal of Environmental Economics and Management, vol. 47, pp. 270-283.

Emam K., Goldenson D., McCurley J. and Herbsleb J., (1998), "Success or failure? Modeling the likelihood of software process improvement", International software engineering research network, Technical Report ISERN-98-15.

Environmental Protection Agency (EPA), Council for Regulatory Environmental Modeling (CREM), "Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models", http://www.epa.gov/osp/crem/library/CREM%20Guidance%20Draft%2012_03.pdf.

European Commission (2000), Business Climate Indicator, DG ECFIN, European Commission, Brussels.

European Commission (2001a), Summary Innovation Index, DG Enterprise, European Commission, Brussels.

European Commission (2001b), Internal Market Scoreboard, DG MARKT, European Commission, Brussels.

European Commission (2004a), Economic Sentiment Indicator, DG ECFIN, Brussels, http://europa.eu.int/comm/economy_finance/index_en.htm

European Commission (2004 b), Composite Indicator on e-business readiness, DG JRC, European Commission, Brussels.

Everitt, B.S. (1979), "Unresolved Problems in Cluster Analysis," Biometrics, 35, 169 -181.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. and Strahan, E. J. (1999), "Evaluating the use of exploratory factor analysis in psychological research", Psychological Methods, 4: 272-299.

Fagerberg J. (2001), Europe at the crossroads: The challenge from innovation-based growth in the Globalising Learning Economy, B. Lundvall and D. Archibugi eds., Oxford Press.

Feldt, L.S., Woodruffe, D.J. and Salih, F.A. (1987), "Statistical Inference for Coefficient Alpha", Applied Psychological Measurement, 11,1, 93-103.

Forman E.H. (1983), "The analytic hierarchy process as a decision support system", Proceedings of the IEEE Computer society.

Freudenberg, M. (2003), "Composite indicators of country performance: a critical assessment", OECD, Paris.

Funtowicz S.O., Munda G. and Paruccini M. (1990), "The aggregation of environmental data using multicriteria methods", Environmetrics, Vol. 1(4), pp. 353-36.

Funtowicz S.O.and Ravetz J.R. (1990), Uncertainty and quality in science for policy, Kluwer Academic Publishers, Dordrecht.

Gentle, J.E., Härdle, W. and Mori, Y. (Eds.) (2004): *Handbook of Computational Statistics: Concepts and Methods*, Springer

Girardin P., Bockstaller C. and Van der Werf H., (2000), "Assessment of potential impacts of agricultural practives on the environment: the AGRO*ECO method", Environmental impact Assessment Review, vol.20, pp. 227-239.

Gorsuch, R. L. (1983), Factor Analysis. Hillsdale, NJ: Lawrence Erlbaum. Orig. ed. 1974.

Gough C., Castells, N. and Funtowicz S., (1998), "Integrated Assessment: an emerging methodology for complex issues", Environmental Modelling and Assessment, n.3, 19-29.

Green, S.B., Lissitz, R.W. and Mulaik, S.A.(1977), "Limitations of coefficient alpha as an index of test unidimensionality", Educational and Psychological Measurement, 37, 827-838.

Green P.E., and Srinivasan V., (1978), "Conjoint analysis in consumer research: issues and outlook", Journal of Consumer Research 5, 103-123.

Golub, Gene H. & van der Vorst, Henk A. (2000): Eigenvalue computation in the 20th century, Journal of Computational and Applied Mathematics, Vol. 123, Iss. 1-2.

Grubb D., and Wells W. (1993), "Employment regulation and patterns of work in EC countries", OECD Economic Studies, n. 21 Winter, 7-58, Paris.

Hair J.F., Anderson R.E., Tatham R.L. and Black W.C. (1995), Multivariate data analysis with readings, fourth ed. Prentice Hall, Englewood Cliffs, NJ.

Hartigan, J.A. (1975), Clustering Algorithms, New York: John Wiley & Sons, Inc.

Harvey A. (1989), Forecasting, structural time series models and the Kalman filter, Cambridge University Press, Cambridge UK.

Hatcher, L. (1994), A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. Cary, NC: SAS Institute. Focus on the CALIS procedure.

Hattie, J. (1985), "Methodology Review: Assessing unidimensionality of tests and items", Applied Psychological Measurement, 9, 2, 139-164.

Hollenstein, H. (1996), "A Composite Indicator of a Firm's Innovativeness. An Empirical Analysis Based on Survey Data for Swiss Manufacturing", Research Policy, 25, 633-45.

Hollenstein, H. (2003), "Innovation Modes in the Swiss Service Sector. A Cluster Analysis Based on Firm-level Data", Research Policy, 32(5), 845-863.

Homma, T. and Saltelli, A. (1996), "Importance measures in global sensitivity analysis of model output", Reliability Engineering and System Safety, 52(1), 1-17.

Hutcheson, G., and Sofroniou N.(1999), The multivariate social scientist: Introductory statistics using generalized linear models, Thousand Oaks, CA: Sage Publications.

Jacobs, R. P. Smith and M. Goddard (2004), "Measuring performance: an examination of composite performance indicators, Centre for Health Economics", Technical Paper Series 29.

Jae-On K., and Mueller C.W. (1978a), Introduction to factor analysis: What it is and how to do it. Thousand Oaks, CA: Sage Publications, Quantitative Applications in the Social Sciences Series, No. 13.

Jae-On K., and Mueller C.W. (1978b), Factor Analysis: Statistical methods and practical issues. Thousand Oaks, CA: Sage Publications, Quantitative Applications in the Social Sciences Series, No. 14.

Jamison, D. and Sandbu, M. (2001), "WHO ranking of health system performance", Science, 293, 1595-1596.

Jencks, S.F., Huff, E.D. and Cuerdon, T. (2003), "Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001", Journal of the American Medical Association, 289(3): 305-12.

Kahma N. (2000), "Measuring environmental quality: an index of pollution", Ecological Economics, vol. 35 pp. 191-202.

Kahn J.R. and Maynard P. (1995), "Conjoint Analysis as a Method of Measuring Use and Non-Use Values of Environmental Goods", paper presented at the American Economic Association.

Kahn J.R. (1998), "Methods for aggregating performance indicators", mimeo, University of Tennessee.

Kahnna N. (2000), "Measuring Environmental quality: an index of pollution", Ecological Economics, 32, 191-202.

Karlsson J. (1998), A systematic approach for prioritizing software requirements, PhD. Dissertation n. 526, Linkoping, Sverige.

Kaufmann D., Kraay A. and Zoido-Lobatón P. (1999), Aggregating Governance Indicators, Policy Research Working Papers, World Bank, http://www.worldbank.org/wbi/governance/working_papers.html

Kaufmann D., Kraay A. and Zoido-Lobatón P. (2003), "Governance matters III: governance Indicators for 1996-2002", mimeo, World Bank.

Keeney R. and Raiffa H. (1976), Decision with multiple objectives: preferences and value trade-offs, Wiley, New York.

Keynes, J. M. (1891), The Scope and Method of Political Economy. London: Macmillan.

King's Fund (2001), The sick list 2000, the NHS from best to worst, http://www.fulcrumtv.com/sick%20list.htm

Kline, R.B. (1998), Principles and practice of structural equation modelling, NY: Guilford Press. Covers confirmatory factor analysis using SEM techniques. See esp. Ch. 7.

Koedijk K. and Kremers J. (1996), "Market opening, regulation and growth in Europe", Economic Policy (0)23, October.

Korhonen P., Tainio R. and Wallenius J., (2001), "Value efficiency analysis of academic research", European Journal of Operational Research, 130, 121-132.

Krantz D.H., Luce R.D., Suppes P. and Tversky A. (1971), Foundations of measurement, vol. 1, Additive and polynomial representations, Academic Press, New York.

Lawley, D. N. and Maxwell A. E. (1971), Factor analysis as a statistical method, London: Butterworth and Co.

Levine, M.S. (1977), Canonical analysis and factor comparison, Thousand Oaks, CA: Sage Publications, Quantitative Applications in the Social Sciences Series, No. 6.

Little R.J.A. and Schenker N. (1994), Missing Data, in Handbook for Statistical Modeling in the Social and Behavioral Sciences (G. Arminger, C.C Clogg, and M.E. Sobel eds.) pp.39-75, New York: Plenum.

Little R.J.A (1997), Biostatistical Analysis with Missing Data, in Encyclopedia of Biostatistics (p. Armitage and T. Colton eds.) London: Wiley.

Little R.J.A. and Rubin D.B. (2002), Statistical Analysis with Missing Data, Wiley Interscience, J. Wiley &Sons, Hoboken, New Jersey.

Mahlberg B. and Obersteiner M. (2001), Remeasuring the HDI by data Envelopment analysis, Interim report IR-01-069, International Institute for Applied System Analysis, Laxenburg, Austria.

Manly B. (1994), Multivariate statistical methods, Chapman & Hall, UK.

Massam B.H. (2002), "Quality of life: public planning and private living", Progress in Planning, vol. 58(3), pp. 141-227.

Massart, D.L. and Kaufman, L. (1983), The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis, New York: John Wiley & Sons, Inc.

McDaniel, C. and Gates R. (1998), Contemporary Marketing Research. West Publishing, Cincinnati, OH.

Melyn W. and Moesen W.W. (1991), "Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available", Public Economic research Paper 17, CES, KU Leuven.

Miller, M.B. (1995), "Coefficient Alpha: a basic introduction from the perspectives of classical test theory and structural equation modelling", Structural Equation Modelling, 2, 3, 255-273.

Milligan, G.W. and Cooper, M.C. (1985), "An Examination of Procedures for Determining the Number of Clusters in a Data Set," Psychometrika, 50, 159 -179.

Moldan, B., Billharz, S. and Matravers, R. (1997), Sustainability Indicators: Report of the Project on Indicators of Sustainable Development, SCOPE 58. Chichester and New York: John Wiley & Sons.

Moldan B., Hak T., Kovanda J., Havranek M. and Kuskova P. (2004), "Composite Indicators of Environmental Sustainability", OECD World Forum on Key Indicators, Palermo, 10-13 November 2004, proceedings, see http://www.oecd.org/dataoecd/43/48/33829383.doc

Muldur U. (2001), Technical annex on structural indicators. Two composite indicators to assess the progress of member States in their transition towards a knowledge based economy, DG RTD, Brussels.

Munda G. (1993), "Fuzzy information in multicriteria environmental evaluation models", PhD dissertation Vrije Universiteit te Amsterdam.

Munda G. (1995), Multicriteria evaluation in a fuzzy environment, Physica-Verlag, Contributions to Economics Series, Heidelberg.

Munda G. (2004), MCDA and Sustainability Decisions, forthcoming in J. Figueira, S. Greco and M. Ehrgott (eds.) – State of the art of multiple-criteria decision analysis, Kluwer, Dordrecht.

Munda, G. and Nardo, M. (2003), "On the methodological foundations of composite indicators used for ranking countries", OECD/JRC Workshop on Composite Indicators of Country Performance, Ispra, Italy, May 12.

Munda, G. and Nardo M. (2003), "On the Construction of Composite Indicators for Ranking Countries", mimeo, Universitat Autonoma de Barcelona.

Nanduri M., Nyboer J. and Jaccard M. (2002), "Aggregating physical intensity indicators: results of applying the composite indicator approach to the Canadian industrial sector", Energy Policy, 30, 151-162.

Nardo, M., Tarantola S., Saltelli A., Andropoulos C., Buescher R., Karageorgos G., Latvala A., and Noel F. (2004), "The e-business readiness composite indicator for 2003: a pilot study", EUR 21294.

Nicoletti G., S. Scarpetta and O. Boylaud, (2000), "Summary indicators of product market regulation with an extension to employment protection legislation", OECD, Economics department working papers No. 226, ECO/WKP(99)18. http://www.oecd.org/eco/eco.

Nilsson R. (2000), "Confidence Indicators and Composite Indicators", CIRET conference, Paris, 10-14 October 2000

NISTEP (National Institute of Science and Technology Policy), (1995), Science and Technology Indicators, NISTEP Report No. 37, Japan.

Norman G. R. and Streiner D. L. (1994), Biostatistics: The bare essentials. St. Louis, MO: Mosby.

Nunnaly J. (1978), Psychometric theory, New York: McGraw-Hill.

OECD (1999), Employment Outlook, Paris.

OECD (2003), Quality Framework and Guidelines for OECD Statistical Activities, www.oecd.org/statistics.

OECD (2004), Learning for Tomorrow's World - First Results from PISA 2003, Programme for International Student Assessment, http://www.pisa.oecd.org/dataoecd/1/60/34002216.pdf

Parker J. (1991), "Environmental reporting and environmental indices", PhD Dissertation, Cambridge, UK.

Pett M.A., Lackey N.R. and Sullivan J.J. (2003), Making sense of factor analysis: The use of factor analysis for instrument development in health care research. Thousand Oaks, CA: Sage Publications.

Pré Consultants (2000), The Eco-indicator 99. A damage oriented method for life cycle impact assessment. http://www.pre.nl/eco-indicator99/ei99-reports.htm

Podinovskii V.V. (1994), "Criteria importance theory", Mathematical Social Sciences, 27, pp. 237 - 252.

Porter M. and Stern S. (1999), The new challenge to America's prosperity: finding from the Innovation Index, Council on Competitiveness, Washington D.C.

Puolamaa M., Kaplas M., and Reinikainen T. (1996), Index of Environmental Friendliness. A methodological study, Eurostat.

Raykov T. (1998b), "Cronbach's Alpha and Reliability of Composite with Interrelated Non-homogenous Items", Applied Psychological Measurement, 22, 375-385.

Rosen R. (1991), Life Itself: A Comprehensive Inquiry into Nature, Origin, and Fabrication of Life. Columbia University Press 1991.

Roubens M. and Vincke P. (1985), Preference Modelling, Springer Verlag, Heidelberg.

Roy B. (1996), Multicriteria methodology for decision analysis, Kluwer, Dordrecht.

Saaty T. L. (1980), The Analytic Hierarchy Process, New York: McGraw-Hill.

Saaty R.W. (1987), "The analytic hierarchy process: what it is and how it is used", Mathematical Modelling, 9, 161-176.

Saisana M. and Tarantola S. (2002), State-of-the-art report on current methodologies and practices for composite indicator development, EUR 20408 EN, European Commission-JRC: Italy.

Saisana M., Tarantola S. and Saltelli A. (2005a), "Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators", Journal of the Royal Statistical Society A, 168(2), 1-17.

Saisana M., Nardo M. and Saltelli A. (2005b) Uncertainty and Sensitivity Analysis of the 2005 Environmental Sustainability Index, in Esty D. Levy M., Srebotnjak T. and de Sherbinin A. 2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship. New Have: Yale Center for Environmental Law and Policy, p.75-78.

Saltelli A. Nardo M., Saisana M. and Tarantola S. (2004) Composite Indicators - The Controversy and the Way Forward, OECD World Forum on Key Indicators, Palermo, 10-13 November 2004, http://www.oecd.org/dataoecd/40/50/33841312.doc

Saltelli A. (2002)," Making best use of model valuations to compute sensitivity indices", Computer Physics Communications, 145, 280-297.

Saltelli A., Chan, K. and Scott M. (2000a), Sensitivity analysis, Probability and Statistics series, New York: John Wiley & Sons.

Saltelli A. and Tarantola S. (2002),"On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal", Journal of American Statistical Association, 97 (459), 702-709.

Saltelli A., Tarantola S. and Campolongo F. (2000b), "Sensitivity analysis as an ingredient of modelling", Statistical Science, 15, 377-395.

Saltelli A., Tarantola S., Campolongo F. and Ratto M. (2004), Sensitivity Analysis in practice, a guide to assessing scientific models, New York: John Wiley & Sons. A software for sensitivity analysis is available at http://www.jrc.cec.eu.int/uasa/prj-sa-soft.asp.

Sharpe, A. (2004), Literature Review of Frameworks for Macro-indicators, Centre for the Study of Living Standards, Ottawa, CAN.

Schumpeter J.A. (1933), "The common sense of econometrics", Econometrica 1: 5-12.

Sobol' I. M. (1993)," Sensitivity analysis for non-linear mathematical models", Mathematical Modelling & Computational Experiment 1, 407-414.

Sobol' I. M. (1967), "On the distribution of points in a cube and the approximate evaluation of integrals", USSR Computational Mathematics and Physics, 7, 86–112.

Spath H. (1980), Cluster Analysis Algorithms, Chichester, England: Ellis Horwood.

SPRG (2001), Report of the Scientific Peer Review Group on Health Systems Performance Assessment, Scientific Peer Review Group (SPRG), WHO: Geneva. http://www.who.int/health-systemsperformance/sprg/report_of_sprg_on_hspa.htm

Storrie D. and Bjurek H. (1999), "Benchmarking European labour market performance with efficiency frontier technique", Discussion Paper FS I 00-2011.

Storrie D. and Bjurek H. (2000), Benchmarking the basic performance indicators using efficiency frontier techniques, Report presented to the European commission, DG employment and social affairs.

Tarantola S., Jesinghaus J. and Puolamaa M. (2000), Global sensitivity analysis: a quality assurance tool in environmental policy modelling. In Sensitivity Analysis (eds Saltelli A., Chan K., Scott M.) pp. 385-397. New York: John Wiley & Sons.

Tarantola S., Saisana M., Saltelli A., Schmiedel F. and Leapman, N. (2002), Statistical techniques and participatory approaches for the composition of the European Internal Market Index 1992-2001, EUR 20547 EN, European Commission: JRC-Italy.

Tarantola S., Liska R., Saltelli A., Leapman N. and Grant C. (2004), The Internal Market Index 2004, EUR 21274EN, European Commission: JRC-Italy

Ting H.M. (1971), Aggregation of attributes for multiattributed utility assessment, Technical report n. 66, Operations Research Center, MIT Cambridge Mass.

Ülengin B., Ülengin F. and Güvenç Ü., (2001), "A multidimensional approach to urban quality of life: the case of Istanbul", European Journal of Operational Research, 130, 361-374.

United Nations (1992, 1999, 2000, 2001), Human Development Report. United Kingdom: Oxford University Press. http://www.undp.org

U.K. Government (2004), "Sustainable Development Indicators in Your Pocket 2004".

U.S. Department of Energy and Energy Information Administration (1995), Measuring energy efficiency un the United States' economy: a beginning. U.S. Department of Energy, Washington, DC.

Vichi M. and Kiers H. (2001), "Factorial k-means analysis for two-way data", Computational Statistics and Data Analysis, 37(1), 49-64.

Vincke Ph. (1992), Multicriteria decision aid, Wiley, New York.

Widaman K. F. (1993), "Common factor analysis versus principal components analysis: Differential bias in representing model parameters?", Multivariate Behavioral Research 28: 263-311. Cited with regard to preference for PFA over PCA in confirmatory factor analysis in SEM.

World Economic Forum (2002), Environmental Sustainability Index. http://www.ciesin_org/indicators/ESI/index.html.

World Economic Forum (2004), Global Competitiveness Report.

WHO (2000), Overall Health System attainment. http://www.who.int/whr2001/2001/archives/2000/en/contents.htm

Young H.P. and Levenglick A. (1978), "A consistent extension of Condorcet's election principle", SIAM Journal of Applied Mathematics, 35, pp. 285-300.

Young H.P. (1988), "Condorcet's theory of voting", American Political Science Review, Vol. 82, No. 4, pp. 1231-1244.

Zimmermann H.J. and Zysno P. (1983), "Decisions and evaluations by hierarchical aggregation of information", Fuzzy Sets and Systems, 10, pp.243-260.

APPENDIX: TECHNOLOGY ACHIEVEMENT INDEX

Table A.1. List of sub-indicators of the Technology Achievement Index

Indicator	Unit	Definition
CREATION OF TECHNOLOGY		
PATENTS	Patents granted per 1,000,000 people	Number of patents granted to residents, to reflect the current level of invention activities (1998)
ROYALTIES	US $ per 1,000 people	Receipts of royalty and license fees from abroad per capita, so as to reflect the stock of successful innovations of the past that are still useful and hence have market value (1999)
DIFFUSION OF RECENT INNOVATIONS		
INTERNET	Internet hosts per 1,000 people	Diffusion of the Internet, which is indispensable to participation in the network age (2000)
EXPORTS	%	Exports of high and medium technology products as a share of total goods exports (1999)
DIFFUSION OF OLD INNOVATIONS		
TELEPHONES	Telephone lines per 1,000 people (log)	Number of telephone lines (mainline and cellular), which represents old innovation needed to use newer technologies and is also pervasive input to a multitude of human activities (1999)
ELECTRICITY	kWh per capita (log)	Electricity consumption, which represents old innovation needed to use newer technologies and is also pervasive input to a multitude of human activities (1998)
HUMAN SKILLS		
SCHOOLING	Years	Mean years of schooling (age 15 and above), which represents the basic education needed to develop cognitive skills (2000)
ENROLMENT	%	Gross enrolment ratio of tertiary students enrolled in science, mathematics and engineering, which reflects the human skills needed to create and absorb innovations (1995-1997)

Table A.2. Raw data for the sub-indicators of the Technology Achievement Index

		PATENTS	ROYALTIES	INTERNET	EXPORTS	TELEPHONES (log)	ELECTRICITY (log)	SCHOOLING	ENROLMENT
1	Finland	187	125.6	200.2	50.7	3.08	4.15	10	27.4
2	United States	289	130	179.1	66.2	3.00	4.07	12	13.9
3	Sweden	271	156.6	125.8	59.7	3.10	4.14	11.4	15.3
4	Japan	994	64.6	49	80.8	3.00	3.86	9.5	10
5	Korea, Rep. of	779	9.8	4.8	66.7	2.97	3.65	10.8	23.2
6	Netherlands	189	151.2	136	50.9	3.02	3.77	9.4	9.5
7	United Kingdom	82	134	57.4	61.9	3.02	3.73	9.4	14.9
8	Canada	31	38.6	108	48.7	2.94	4.18	11.6	14.2
9	Australia	75	18.2	125.9	16.2	2.94	3.94	10.9	25.3
10	Singapore	8	25.5	72.3	74.9	2.95	3.83	7.1	24.2
11	Germany	235	36.8	41.2	64.2	2.94	3.75	10.2	14.4
12	Norway	103	20.2	193.6	19	3.12	4.39	11.9	11.2
13	Ireland	106	110.3	48.6	53.6	2.97	3.68	9.4	12.3
14	Belgium	72	73.9	58.9	47.6	2.91	3.86	9.3	13.6
15	New Zealand	103	13	146.7	15.4	2.86	3.91	11.7	13.1
16	Austria	165	14.8	84.2	50.3	2.99	3.79	8.4	13.6
17	France	205	33.6	36.4	58.9	2.97	3.80	7.9	12.6
18	Israel	74	43.6	43.2	45	2.96	3.74	9.6	11
19	Spain	42	8.6	21	53.4	2.86	3.62	7.3	15.6
20	Italy	13	9.8	30.4	51	3.00	3.65	7.2	13
21	Czech Republic	28	4.2	25	51.7	2.75	3.68	9.5	8.2
22	Hungary	26	6.2	21.6	63.5	2.73	3.46	9.1	7.7
23	Slovenia	105	4	20.3	49.5	2.84	3.71	7.1	10.6
24	Hong Kong, China (SAR)	6		33.6	33.6	3.08	3.72	9.4	9.8
25	Slovakia	24	2.7	10.2	48.7	2.68	3.59	9.3	9.5
26	Greece			16.4	17.9	2.92	3.57	8.7	17.2
27	Portugal	6	2.7	17.7	40.7	2.95	3.53	5.9	12
28	Bulgaria	23		3.7	30	2.60	3.50	9.5	10.3
29	Poland	30	0.6	11.4	36.2	2.56	3.39	9.8	6.6
30	Malaysia			2.4	67.4	2.53	3.41	6.8	3.3
31	Croatia	9		6.7	41.7	2.63	3.39	6.3	10.6
32	Mexico	1	0.4	9.2	66.3	2.28	3.18	7.2	5
33	Cyprus			16.9	23	2.87	3.54	9.2	4
34	Argentina	8	0.5	8.7	19	2.51	3.28	8.8	12
35	Romania	71	0.2	2.7	25.3	2.36	3.21	9.5	7.2
36	Costa Rica		0.3	4.1	52.6	2.38	3.16	6.1	5.7
37	Chile		6.6	6.2	6.1	2.55	3.32	7.6	13.2
38	Uruguay	2		19.6	13.3	2.56	3.25	7.6	7.3
39	South Africa		1.7	8.4	30.2	2.43	3.58	6.1	3.4
40	Thailand	1	0.3	1.6	48.9	2.09	3.13	6.5	4.6
41	Trinidad and Tobago			7.7	14.2	2.39	3.54	7.8	3.3
42	Panama			1.9	5.1	2.40	3.08	8.6	8.5

43	Brazil	2	0.8	7.2	32.9	2.38	3.25	4.9	3.4
44	Philippines		0.1	0.4	32.8	1.89	2.65	8.2	5.2
45	China	1	0.1	0.1	39	2.08	2.87	6.4	3.2
46	Bolivia	1	0.2	0.3	26	2.05	2.61	5.6	7.7
47	Colombia	1	0.2	1.9	13.7	2.37	2.94	5.3	5.2
48	Peru		0.2	0.7	2.9	2.03	2.81	7.6	7.5
49	Jamaica		2.4	0.4	1.5	2.41	3.35	5.3	1.6
50	Iran, Islamic Rep. of	1			2	2.12	3.13	5.3	6.5
51	Tunisia		1.1		19.7	1.98	2.92	5	3.8
52	Paraguay		35.3	0.5	2	2.14	2.88	6.2	2.2
53	Ecuador			0.3	3.2	2.09	2.80	6.4	6
54	El Salvador		0.2	0.3	19.2	2.14	2.75	5.2	3.6
55	Dominican Republic			1.7	5.7	2.17	2.80	4.9	5.7
56	Syrian Arab Republic				1.2	2.01	2.92	5.8	4.6
57	Egypt		0.7	0.1	8.8	1.89	2.94	5.5	2.9
58	Algeria				1	1.73	2.75	5.4	6
59	Zimbabwe			0.5	12	1.56	2.95	5.4	1.6
60	Indonesia			0.2	17.9	1.60	2.51	5	3.1
61	Honduras				8.2	1.76	2.65	4.8	3
62	Sri Lanka			0.2	5.2	1.69	2.39	6.9	1.4
63	India	1		0.1	16.6	1.45	2.58	5.1	1.7
64	Nicaragua			0.4	3.6	1.59	2.45	4.6	3.8
65	Pakistan			0.1	7.9	1.38	2.53	3.9	1.4
66	Senegal			0.2	28.5	1.43	2.05	2.6	0.5
67	Ghana				4.1	1.08	2.46	3.9	0.4
68	Kenya			0.2	7.2	1.04	2.11	4.2	0.3
69	Nepal			0.1	1.9	1.08	1.67	2.4	0.7
70	Tanzania, U. Rep. of				6.7	0.78	1.73	2.7	0.2
71	Sudan				0.4	0.95	1.67	2.1	0.7
72	Mozambique				12.2	0.70	1.73	1.1	0.2

Note: The first 23 countries are used as case study in the Handbook. Units are given in Table A.1.

ENDNOTES

[1] The technique of PCA was first described by Karl Pearson in 1901. A description of practical computing methods came much later from Hotelling in 1933. For a detailed discussion on the PCA, *see* Jolliffe, (1986), Jackson (1991) and Manly (1994). Social scientists may also find the shorter monograph by Dunteman, (1989) to be helpful.

[2] For reasons of clarity in this section we substitute the indexing $q=1,\ldots Q$ with the indexing $i=1,\ldots,Q$ and $j=1,\ldots,Q$.

[3] Golub, Gene H. & van der Vorst, Henk A. (2000): *Eigenvalue computation in the 20th century*, Journal of Computational and Applied Mathematics, Vol. 123, Iss. 1-2. And Gentle, James E.; Härdle, Wolfgang; Mori, Yuichi (Eds.) (2004): *Handbook of Computational Statistics: Concepts and Methods*, Springer

[4] Euclidean distances can be greatly influenced by variables that have the largest values. One way around this problem is to standardise the variables.

[5] The name is based on the route that follows the grid of roads, as in most American cities it is not possible to go directly between two points.

[6] The value that divides in two equal parts the distribution of the random variable

[7] The value with the highest frequency

[8] A variant of unconditional mean imputation is the fill-in via conditional mean. The regression approach is one possible method. Another common method (called imputing means within adjustment cells) is to classify the data for the sub-indicator with some missing values in classes and impute provisionally the missing values of that class with the sample mean of the class. Then sample mean (across al classes) is then calculated and substituted as final imputation value.

[9] If the observed variables are dummies for a categorical variable then the prediction (7*) are respondent means within classes defined by the variable and the method reduces to that of imputing means with adjustment cells.

[10] Define $SSE = \sum_i (x_{ih} - \hat{x}_{ih})^2$, $SST = \sum_i (x_{ih} - \bar{x}_h)^2$, then $R^2 = 1 - (SSE / SST)$, $MSE = SSE / (M - r - k)$, where k is the number of coefficients in the regression and *(M-r)* the number of observations. $RMS = \sum_i (\hat{x}_{ih} - \bar{x}_h)^2$ and $C_k = (SSE_k / MSE) - (M - r) + 2k$ where the SSE_k is computed from a model with only k coefficients and MSE is computed using all available regressors.

[11] Other iterative methods include the Newton-Raphson algorithm and the scoring method. Both involve a calculation of the matrix of second derivatives of the likelihood, which, for complex pattern of incomplete data, can be a very complicate function of θ. As a result these algorithms often require algebraic manipulations and complex programming. Numerical estimation of this matrix is also possible but careful computation is needed.

[12] For NMAR mechanisms one needs to make assumption on the missing-data mechanism and include them into the model, see Little and Rubin, (2002), Ch. 15.

[13] In a sample of *n* observations it is possible for a limited number to be so far separated in value from the remainder that they give rise to the question whether they are not from a different population, or that the sampling technique is a fault. Such values are called outliers (F.H.C. Marriott, 1990, A dictionary of statistical terms, Longman Scientific & Technical, Fifth edition, p.223). Eurostat adopts this definition of outlier.

[14] http://europa.eu.int/comm/economy_finance/publications/european_economy/2001/b2001_0809_en.pdf)

[15] Other methods are available, e.g. the Maximum Likelihood or the principal Factor centroids. Notice that these methods usually supply very different weights especially when the sample size of FA is small.

[16] Weights are normalized squared factor loading, e.g. 0.24 = (0.79^2)/2.64 which is the portion of the variance of the first factor explained by the variable Internet.

[17] To preserve comparability final weights could be rescaled to sum up to one.

[18] DEA has also been used in production theory, for a review see Charnes et al., (1995).

[19] We present the method as it has been used in Cherchye et al., (2004), and Cherchye and Kuosmanen, (2002).

[20] Additional constraints could be imposed. Country-specific restrictions to reflect prior information can also be added. Notice that constraints should not be given to the absolute value of weights but placed on the relative weights. The result of this approach is therefore relative weights or trade offs to be legitimately used in linear and geometric aggregations.

[21] In our example we imposed the requirement for each sub-indicator to weight at least 10% and no more than 15% of the total.

[22] However, precisely since BOD weights have the meaning of relative weights, the weights as originally produced by the algorithm could be normalised afterwards so as to sum up to one facilitating the comparison with the results of other methods.

[23] In 1991, 400 German experts were asked to allocate a budget to several environmental indicators related to an air pollution problem. The results were consistent, although the experts came from opposing social spheres like the industrial and the environmental sectors (Jesinghaus in Moldan and Billharz, 1997).

[24] The exercise was carried out at JRC interviewing experts in the field.

[25] A subset of indicators *Y* is *preferentially independent* of Y^C (the complement of *Y*) only if any conditional preference among elements of *Y*, holding all elements of Y^C fixed, remain the same, regardless of the levels at which Y^C are held. The variables $x_1, x_2,..., x_Q$ are *mutually preferentially independent* if every subset *Y* of these variables is preferentially independent of its complementary set of evaluators.

[26] $\dfrac{\partial S_{x,y}}{\partial z} = 0$, $\forall x, y \in Y, \forall z \in Y^C$, see the previous note above.

[27] Suppose that country *a* is evaluated according to some criteria/sub-indicators $(x_1(a),...,x_Q(a))$, then the *substitution rate at a*, of sub-indicator *j* with respect to sub-indicator *r* (taken as a reference) is the amount $S_{jr}(a)$ such that, country *b* whose evaluations are: $x_l(a) = x_l(b), \forall l \neq j, r$; $x_j(b) = x_j(a) - 1$;

and $x_r(b) = x_r(a) + S_{jr}(a)$ is indifferent to country *a*. Therefore, $S_{jr}(a)$ is the amount which must be added to the reference sub-indicator in order to compensate the loss of one unit on sub-indicator *j* keeping constant the others. While for additive aggregations the substitution rate is constant, in the multiplicative aggregation it is proportional to the relative score of the indicator with respect to the others.

[28] Data are not normalized. Normalization does not change the result of the multicriteria method whenever it does not change the ordinal information of the data matrix.

[29] To prevent this problem it is possible to set thresholds of this type: if the difference between two countries in the indicator I is more than x%, then give to the country with the highest score a much higher weight. If the difference is less than x% give nearly the same weight. However, more precision comes at the expenses of ad hoc threshold and weighting values.

[30] Examples of MCA include the Environmental Sustainability Indicator 2005, agro-ecological indicators (Girardin et al., 2000) and an indicator of quality of life of three towns near to Puerto Vallarta, Mexico (Massam, 2002) within a project on the effects of tourism on the quality of life of small communities near international tourist resorts.

[31] Compensability of aggregations is widely studied in fuzzy sets theory, for example Zimmermann and Zysno, (1983) use the geometric operator $(\prod_q I_q)^{(1-\gamma)}(1-\prod_q(1-I_q))^{\gamma}$ where γ is a parameter of compensation: the larger is γ the higher is the degree of compensation between operators (in our case sub-indicators).

[32] The multi-criteria approach MCA produces ranks for countries. The focus of the analysis thus is $Rank(CI_c)$ rather than the raw value of the index CI_c.

[33] For proof, see Saltelli et al., (2004).