8-2008

Sustainability Evaluation: Development and Validation of an Evaluation Checklist

Daniela C. Schroeter
Western Michigan University

Follow this and additional works at: http://scholarworks.wmich.edu/dissertations
Part of the Environmental Policy Commons, and the Public Administration Commons

Recommended Citation

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
SUSTAINABILITY EVALUATION: DEVELOPMENT AND VALIDATION OF AN EVALUATION CHECKLIST

by

Daniela C. Schroeter

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Interdisciplinary Ph.D. in Evaluation
Dr. Matthew Mingus and Dr. Michael Scriven, Advisors

Western Michigan University
Kalamazoo, Michigan
August 2008
ACKNOWLEDGEMENTS

Es ist nicht genug, zu wissen, man muß auch anwenden; es ist nicht genug, zu wollen, man muß auch tun.
— Knowing is not enough, you must apply; willingness is not enough, you must do.
—Johann Wolfgang von Goethe

Ich, der ich weiß, mir einzubilden, dass ich weiß, nichts zu wissen, weiß, dass ich nichts weiß.
— I, who know that I imagine, that I know to know nothing, know that I don't know.
—Socrates

Much of my dissertation experience can be summarized by knowing and not knowing, doing and not doing, optimism and doubts, and in the end, knowing to know nothing, but still finally climbing the mountain, called dissertation defense! The quotes presented above describe this experience well, and I am lucky that I got to know all of you who wholeheartedly supported me in this endeavor. No thank you note can ever express what should be said, yet here I go:

I would like to begin with my dissertation committee, my doctoral fathers, Dr. Matthew S. Mingus (Chair) and Dr. Michael J. Scriven (Director), and my doctoral mother, Dr. Jennifer Palthe (Member) and express my appreciation for their confidence in me to pursue this work. Dr. Mingus’s leadership and engagement enabled my dissertation defense to happen now, rather than later. Thank you for taking over the steering wheel, facilitating my Germanic quest for order, and pushing me to this day. Dr. Scriven—my vigorous mentor who dedicated countless hours to mentoring me over the years and from whom I have learned more than anyone in my
pursuit to become an evaluator, thank you for all the brilliance you have brought to this doctoral program and for taking the time to provide exquisite advise on my dissertation at your home. With this, I would also like to thank your wife, Mary Ann, for welcoming my husband and me to your home. My gratitude extends to Dr. Palthe, who has a unique ability to motivate students to expand their horizons, to collaborate in the true sense of the word, and to recognize and realize their potential. Thank you for your exuberance!

My gratitude goes to my program committee: Dr. E. Jane Davidson (later Dr. Michael Scriven), Dr. Daniel L. Stufflebeam, and Dr. Brooks Applegate. Dr. Davidson, my first advisor and mentor, accepted me into the program and taught me the nuts and bolts of evaluation. Thank you, Dr. Stufflebeam, from whom I learned much about the history of evaluation and who provided a lot of guidance through the course of my study. Thank you, Dr. Applegate, my favorite statistician, who encouraged me to go beyond the basic statistics coursework to complete a course in advanced multivariate statistics.

Much appreciation goes to the current and former staff of Western Michigan University's Evaluation Center. Dr. Arlen Gullickson took me on as one of the research associates on the evaluation of the National Science Foundation’s Advanced Technological Education program (NSF ATE). Later we collaborated on the evaluation of Heifer Project International in Kenya, where he also provided much mentorship and guidance. Thank you, Arlen. I had an opportunity to work with Dr.
Acknowledgements—continued

Gary Miron on the Heifer Project International Evaluation in Nepal and in conducting field work for educational evaluations in Hawaii. He also supervised my project with the German Youth Institute and provided much guidance on managerial and administrative tasks associated with the project. Thank you, Gary. My gratitude is also expanded to Anna and your family for their sensitivity and moral support. Dr. James Sanders, thank you for your advice and for sharing your experience and expertise in many hallway conversations, as well as your contribution to improving my checklist. Dr. Steven Magura, thank you for allowing me to use The Evaluation Center equipment during the time of the dissertation. Dr. Carl Hanssen, it was a pleasure working with you on NSF ATE project as well as on the evaluation of the Making It Work program. Dr. Liesel Ritchie, who has become a great friend and colleague and who is sorely missed here in Michigan, thank you for providing me the opportunity to work on the American Evaluation Association’s (AEA) Birth Story project, as well as on the Healthier Communities evaluation. Thank you, Liesel. Dr. Carolyn Sullins taught the first course I took in evaluation and is responsible for me entering the IDPE program. Thank you, Carolyn. I would like to extend my gratitude to Dr. Pamela Zeller, Sally Veeder, Joseph Fee, Christine Hummel, Barbara Wygant, and Patricia Negrevski. Special thanks go to Mary Ramlow with her motherly nature and vigorous support. Without you, I would have never made it.

I would also like to thank the members of The Graduate College at Western Michigan University, especially Dr. Lewis Pyenson (Dean), Dr. Marianne Di Pierro
(Director of the Graduate Center for Research and Retention), and Dr. Jennifer Wendling (Dissertation Coordinator). Thanks also to Victoria Janson for providing timely and invaluable guidance and feedback on the Human Subjects Institutional Review Board protocol.

Many individuals generously provided their time and efforts in revising and critiquing various aspects of this dissertation. Foremost, I would like to thank Anne Cullen and Amy Gullickson, my good friends and peers who kept me sane during the process and provided feedback and all the moral support one could possibly ask for during such a project. Thank you, Anne and Amy. Thanks to Wes Martz, my “dissertation brother,” who went through this experience at the same time and shared many of the highs and lows. My gratitude extends to Lori Wingate, who edited much of this dissertation.

Special thanks go also to Jim Rugh, who led me to significantly improve my questionnaire and to access the right audience from which to gather feedback on the checklist. Thank you, Jim. And thanks to all the individuals who provided extensive feedback on the checklist by participating in the research described in the study as well as those who provided critical feedback on early versions of the checklist. My gratitude extends to Michael Quinn Patton who cordially provided a preprint copy of his forthcoming edition of *Utilization-focused Evaluation*, of which I made extensive use. Thank you, Jonny Morell, for allowing me to bounce ideas off him over the
Acknowledgements—continued

years. Thank you, Oded Löwenbein, for providing the opportunity to publish my initial thoughts as a book chapter.

I would also like to thank all current and former IDPE students, especially the early cohorts who shared much coursework with me and who have influenced my thinking throughout the years: Drs. Chris L. S. Coryn, Thomaz Chianca, John Risley, and Nadini Persaud; Wes Martz, Otto Gustafson, Krystin Martens, Anne Cullen, Amy Gullickson, Brandon Youker, P. Cristian Gugiu, Brad Watts, Paul Lamphear, Lori Wingate, Tererai Trent, Sheri Hudachek-Lantz, Todd Harcek, Stephanie Evergreen, Willis Thomas, Ronald Visscher, Ryo Sasaki, Mina Zadeh, Stephanie Means, Brandy Pratt, Kristin Richardson, Margaret Richardson, Michelle Woodhouse-Jackson, and Ezechukwu Awgu. I have special appreciation for all of you who participated in the pilot study of the checklist.

Thanks to everyone who participated in the 2007 AEA checklist panel, providing critical feedback and insights. Especially, I thank Mary Ann Scheirer, who provided alternative scenarios to my checklist. My gratitude extends to all interviewees and questionnaire participants. You all spent a tremendous amount of time reading the checklist and providing feedback to me. Thank you.

I thank my parents, Thea (in memoriam) and Otto Schröter, without whom I would not have begun a doctoral program and who encouraged me to go to the United States, although they would have preferred that I stay home. This thanks must be expanded to Karin Goedecke, who provided a lot of emotional support during the
time my father was in the hospital in serious conditions. Without her, I would have left the dissertation and went home to be with my father and not be able to defend now. Danke, Papa und Karin.

Above all, I would like to thank Chris, my husband, my best friend, my peer, my colleague, my favorite methodologist, my life—and the new director of the IDPE. Thank you for putting up with me in the last few months!

Daniela C. Schroeter
TABLE OF CONTENTS

ACKNOWLEDGEMENTS........................................................................................................ ii
LIST OF TABLES.................................................................................................................. xiii
LIST OF FIGURES................................................................................................................ xiv
CHAPTER

I. INTRODUCTION............................................................................................................. 1
   Linguistic Derivation of Sustainability................................................................. 1
   Precursors............................................................................................................... 3
   Sustainability and Sustainable Development................................................... 5
   Sustainability Evaluation.................................................................................... 10
   Problem Statement............................................................................................... 14
   Aim and Scope of the Dissertation..................................................................... 15
   Organization of the Dissertation........................................................................ 17

II. SUSTAINABILITY EVALUATION............................................................................. 20
   Definitions............................................................................................................ 24
   Purposes............................................................................................................... 29
   Evaluands........................................................................................................... 31
   Policies................................................................................................................ 31
   Programs and Projects....................................................................................... 34
# Table of Contents—continued

## CHAPTER

<table>
<thead>
<tr>
<th>Subject</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sustainability as Evaluand</td>
<td>36</td>
</tr>
<tr>
<td>Components of Sustainability Evaluands</td>
<td>44</td>
</tr>
<tr>
<td>Stakeholders and Impactees</td>
<td>46</td>
</tr>
<tr>
<td>Dimensions of Sustainability</td>
<td>50</td>
</tr>
<tr>
<td>Logic of Evaluative Reasoning</td>
<td>64</td>
</tr>
<tr>
<td>Values and Criteria</td>
<td>65</td>
</tr>
<tr>
<td>Standards</td>
<td>73</td>
</tr>
<tr>
<td>Measure Performance and Compare with Standards</td>
<td>74</td>
</tr>
<tr>
<td>Synthesis</td>
<td>77</td>
</tr>
<tr>
<td>Summary</td>
<td>78</td>
</tr>
</tbody>
</table>

### III. SUSTAINABILITY EVALUATION CHECKLIST DEVELOPMENT 79

<table>
<thead>
<tr>
<th>Subject</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background for Checklist Development</td>
<td>80</td>
</tr>
<tr>
<td>What are Evaluation Checklists?</td>
<td>80</td>
</tr>
<tr>
<td>Why Develop Evaluation Checklists?</td>
<td>83</td>
</tr>
<tr>
<td>What Other Sustainability Checklists are in Existence?</td>
<td>85</td>
</tr>
<tr>
<td>Checklist Development Methodology</td>
<td>88</td>
</tr>
<tr>
<td>Models for Checklist Development</td>
<td>89</td>
</tr>
<tr>
<td>Considerations for Formatting Checklists</td>
<td>92</td>
</tr>
<tr>
<td>Procedures Used for Developing the SEC</td>
<td>93</td>
</tr>
</tbody>
</table>
# Table of Contents—continued

## CHAPTER

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Sustainability Evaluation Checklist</td>
<td>94</td>
</tr>
<tr>
<td>Key Concepts</td>
<td>95</td>
</tr>
<tr>
<td>Checklist Content</td>
<td>96</td>
</tr>
<tr>
<td>Summary</td>
<td>105</td>
</tr>
<tr>
<td>IV. METHODOLOGY</td>
<td>106</td>
</tr>
<tr>
<td>Purpose</td>
<td>107</td>
</tr>
<tr>
<td>Participants</td>
<td>113</td>
</tr>
<tr>
<td>Sampling Procedures</td>
<td>113</td>
</tr>
<tr>
<td>Recruitment</td>
<td>114</td>
</tr>
<tr>
<td>Protection of Human Subjects</td>
<td>116</td>
</tr>
<tr>
<td>Research Procedure</td>
<td>117</td>
</tr>
<tr>
<td>Methods of Data Collection</td>
<td>118</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>118</td>
</tr>
<tr>
<td>Measures</td>
<td>123</td>
</tr>
<tr>
<td>Pretest, Pilot Test, and Think-aloud</td>
<td>125</td>
</tr>
<tr>
<td>Location of Data</td>
<td>126</td>
</tr>
<tr>
<td>Data-analysis Procedures</td>
<td>126</td>
</tr>
<tr>
<td>V. RESULTS</td>
<td>129</td>
</tr>
<tr>
<td>Respondents</td>
<td>131</td>
</tr>
</tbody>
</table>
## Table of Contents—continued

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire Respondent Characteristics</td>
<td>132</td>
</tr>
<tr>
<td>Characteristics of Interviewees and E-mail Correspondents</td>
<td>144</td>
</tr>
<tr>
<td>SEC Validity Rating Scales</td>
<td>145</td>
</tr>
<tr>
<td>Reliability and Validity of the Accuracy and Utility Subscales</td>
<td></td>
</tr>
<tr>
<td>Checklist Accuracy</td>
<td>147</td>
</tr>
<tr>
<td>Checklist Utility</td>
<td>150</td>
</tr>
<tr>
<td>Other Checklist Characteristics</td>
<td>151</td>
</tr>
<tr>
<td>Strengths, Weaknesses, and Recommendations</td>
<td>155</td>
</tr>
<tr>
<td>Cross-case Analysis</td>
<td>155</td>
</tr>
<tr>
<td>Cross-item Analysis</td>
<td>177</td>
</tr>
<tr>
<td>Cross-section Analysis</td>
<td>180</td>
</tr>
<tr>
<td>Triangulation</td>
<td>181</td>
</tr>
<tr>
<td>Summary and Implications for Checklist Improvement</td>
<td>184</td>
</tr>
<tr>
<td>VI. DISCUSSION AND CONCLUSIONS</td>
<td>186</td>
</tr>
<tr>
<td>Revisions</td>
<td>187</td>
</tr>
<tr>
<td>Changes to Format and Structure</td>
<td>193</td>
</tr>
<tr>
<td>Omissions: That which has not been done</td>
<td>194</td>
</tr>
<tr>
<td>Limitations and Outlook</td>
<td>195</td>
</tr>
<tr>
<td>Contributions to Evaluation Theory and Practice</td>
<td>198</td>
</tr>
</tbody>
</table>
# Table of Contents—continued

## CHAPTER

Concluding Remarks and Future Directions ........................................ 199

## REFERENCES ................................................................. 200

## APPENDICES ............................................................. 222

A. Acronyms ........................................................................ 222

B. Sustainability Evaluation Checklist (Test Version) ...................... 226

C. Expert Recruitment and Consent ........................................... 245

D. Specific Request to Expert .................................................. 247

E. E-mail to Listservs ............................................................ 248

F. HSIRB Notice (Scan) ......................................................... 250

G. Invitation to Survey ........................................................ 251

H. Interview Protocol .......................................................... 253

I. Online Questionnaire ....................................................... 255

K. Interview Summaries ....................................................... 260

L. Sustainability Evaluation Checklist (Revised Version) ............... 293
# LIST OF TABLES

1. Definitions of Evaluation .......................................................... 26
2. Programming Levels ................................................................. 35
3. SD Categories and Characteristics (Based on Hardi, 2007) and Example Definitions ...................................................... 40
4. Components of Sustainability Evaluands ...................................... 45
5. Definitions of Users, Clients, Audiences, Stakeholders, and Consumers .......................................................... 47
6. Alpha-beta Mixes of Technologies (based on Galtung, 1980a) ......... 60
7. A Taxonomy of Checklist Types (based on Scriven, 2007) ............... 82
8. Sustainability Checklists: A Few Examples .................................. 86
9. Accuracy and Utility Scales .......................................................... 124
10. Modes of Communication, Data Types, and Data Storage .......... 127
11. Pre-identified Respondents ......................................................... 131
12. Correlations of Each SEC Validity Items With its Own Scale and the Other Scale (Accuracy or Utility) After Removing Inadequate Items .... 147
13. Overview of Items, Questions, Responses, Codes, and Themes ........ 157
14. Changes to Content and Format/Structure .................................. 188
15. SEC Summary Table: Quick Guide ............................................. 190
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The House of Evaluation</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Relationships Between Different Stakeholder Groups</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>Power, Interest, and Stakeholders Involvement (inspired by and adapted from Patton’s [2008, p. 80] Power Versus Interest Grid)</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>Sustainability Requirements as Decision Criteria (based on Gibson et al., 2005, pp.116-118)</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>General Logic of Evaluation (based on Fournier, 1995, p. 16)</td>
<td>64</td>
</tr>
<tr>
<td>6</td>
<td>Life Cycles and Prospective/Retrospective Evaluation</td>
<td>76</td>
</tr>
<tr>
<td>7</td>
<td>Chapter III Overview</td>
<td>79</td>
</tr>
<tr>
<td>8</td>
<td>Models for Checklist Development</td>
<td>90</td>
</tr>
<tr>
<td>9</td>
<td>Checklist Flow</td>
<td>98</td>
</tr>
<tr>
<td>10</td>
<td>Chapter IV Overview</td>
<td>106</td>
</tr>
<tr>
<td>12</td>
<td>Study Timeline</td>
<td>117</td>
</tr>
<tr>
<td>13</td>
<td>Screen 1 of Web-Based Questionnaire</td>
<td>120</td>
</tr>
<tr>
<td>14</td>
<td>Final Screen of the Web-based Questionnaire</td>
<td>123</td>
</tr>
<tr>
<td>15</td>
<td>Chapter V Overview</td>
<td>129</td>
</tr>
<tr>
<td>16</td>
<td>Organizational Affiliations</td>
<td>133</td>
</tr>
<tr>
<td>17</td>
<td>Sectors</td>
<td>135</td>
</tr>
<tr>
<td>18</td>
<td>Regions Where Respondents Work</td>
<td>136</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>19</td>
<td>Respondents’ Home Continents</td>
<td>137</td>
</tr>
<tr>
<td>20</td>
<td>Respondents’ Education Level</td>
<td>138</td>
</tr>
<tr>
<td>21</td>
<td>Checklist Performance on Accuracy Subscale</td>
<td>149</td>
</tr>
<tr>
<td>22</td>
<td>Frequencies Distribution of Checklist Performance on Accuracy Items</td>
<td>149</td>
</tr>
<tr>
<td>23</td>
<td>Checklist Performance on Utility Subscale</td>
<td>151</td>
</tr>
<tr>
<td>24</td>
<td>Frequencies Distribution of Checklist Performance on Utility Items</td>
<td>152</td>
</tr>
<tr>
<td>25</td>
<td>Overview of Qualitative Analysis</td>
<td>156</td>
</tr>
<tr>
<td>26</td>
<td>Omissions: Frequency of Emergent Themes</td>
<td>158</td>
</tr>
<tr>
<td>27</td>
<td>Commissions: Frequency of Emergent Themes</td>
<td>162</td>
</tr>
<tr>
<td>28</td>
<td>Confusions: Frequency of Emergent Themes</td>
<td>166</td>
</tr>
<tr>
<td>29</td>
<td>Errors and Problems: Frequency of Emergent Themes</td>
<td>169</td>
</tr>
<tr>
<td>30</td>
<td>Strengths: Frequency of Emergent Themes</td>
<td>171</td>
</tr>
<tr>
<td>31</td>
<td>Suggestions: Frequency of Emergent Themes</td>
<td>176</td>
</tr>
<tr>
<td>32</td>
<td>Cross-item Analysis</td>
<td>178</td>
</tr>
<tr>
<td>33</td>
<td>Chapter VI Overview</td>
<td>187</td>
</tr>
<tr>
<td>34</td>
<td>SEC Flow Chart</td>
<td>192</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

Sustainability is a buzzword that permeates many aspects of human activity and has been defined in varying ways. A simple Google search\(^1\) on January 31, 2008, yielded more than 17 million entries of the noun sustainability; 530,000 entries for sustainability evaluation; 7,210 entries for exact wording of “sustainability evaluation”; and 242,000 for “sustainability checklist”; but only 7 entries for “sustainability evaluation checklist.” Refining this search to Google Scholar left 134,000 entries, of which 5,790 were 2007 contributions. Filtering the search to only include English-language contributions limited the results to 130,000 entries, of which 5,580 were published in 2007 alone. Most of these directly relate to global sustainable development issues and fewer to the continuation of projects, programs, or policies.

Linguistic Derivation of Sustainability

Etymologically, the meaning of “sustain” derives from Middle English, Old French, and Latin. Based on Latin, the prefix sus- is a variant of sub-, meaning up. The stem -tain is derived from tinère or tenère, meaning to hold (c.f. Merriam-

\(^{1}\) Note that repeated searches yield slightly different results at the same time, so these numbers only reflect approximations; the more specific the search term, the smaller the changes in repeated searches. Side note: Both Google and Google Scholar will stop continuing searches after several tries assuming that the search is related to a virus or other inappropriate function.
Webster, 2003). Therefore, it can be inferred that sustain means to hold up, keep up, prolong, or continue. As a transitive verb, one that requires a grammatical subject as well as a grammatical object, sustain calls for consideration of (1) who/what enables the action or state, and (2) who/what is impacted by the action or state. As a qualifier of a noun, sustainable relates to something being “capable of being sustained; of, relating to, or being a method of harvesting or using a resource so that the resource is not depleted or permanently damaged; [or] of or relating to a lifestyle involving the use of sustainable methods” (Merriam-Webster, 2003).

The noun sustainability, although widely used, is not yet included in English language dictionaries such as Merriam Webster or the Oxford English Dictionary (Soanes & Stevenson, 2004). Linguistically, it is a composite of to sustain and ability. The noun suffix -ability indicates a certain potential or means to an end, that is, “capacity, fitness, or tendency to act or be acted on in a (specific) way” (Merriam-Webster, 2003). In the evaluation world, for example, people commonly talk about evaluability assessment (c.f., Wholey, 2004; Scriven, 2006), in which characteristics of evaluands are examined for their capacities to allow certain types of evaluation or classes of evaluation design.

In evaluative terms, the verb subject can be thought of as a group of people, an organization, stakeholders, and impactees who continue a given evaluand, that is, an evaluation object (e.g., a program, resources, outcomes from a program, etc.). The questions may be: what (object or objective) needs to be sustained by whom
(subject)? While this question per se is not evaluative because it does not address concerns about merit, worth, or significance (i.e., the key definitional properties of evaluation) of a given evaluation object, it provides the context in which evaluation occurs.

**Precursors**

The idea to develop a sustainability evaluation checklist grew out from initial work on evaluation projects that considered sustainability. The first project was a sustainability evaluation of a national program funded by a large government agency. The second was an international development evaluation. Brief stories from these experiences are included below:

A few years ago, I was involved in an evaluation of the sustainability of a national program funded by a large government agency. In this program, community colleges received grants to develop and disseminate courses, curricula, professional development, materials, articulation agreements, and conduct research to increase the number of competent technicians in the workforce. **Sustainability in this context was perceived as the continuation of the funded work streams after the initial funding for the project had ended.** The information was intended to be useful in justifying the program’s existence and to yield suggestions for improvement of sustainability. The key questions of the evaluation related to continued productivity, collaboration, and support as well as adoption of program activities by other institutions. To answer these questions, former project directors were surveyed. The sample consisted of almost 200 projects, of which about 80 percent responded to the survey. Despite intense telephone follow-up, some informants could not be reached and others chose not to respond. This was not surprising, given that these projects had not received funding for up to nine years. However, for the projects that did respond, the median support was for 4 years and $450,000. Findings from the study suggested that: Productivity was sustained (via institutionalization) for most activities. Collaboration dropped by half over
time, but still showed quite high numbers. Support (esp. monetary support) decreased rapidly although collaborations profit and nonprofit organizations were in place. Adoption took place to some degree, but usage was not confirmed via the evaluation. In essence, although production persisted, collaboration and funding decreased and the degree of adoption was not confirmed. So what did we learn about the merit, worth, or significance of sustainability and sustained projects or the program? While continuation of some program components were identified, the results yielded little information about the quality and use of what was continued; its value to people, organizations, or society; and the general importance of what was done. In essence, we did not learn if the graduates of these programs joined the intended workforce (sustained their education), if the employers perceived these graduates as equally good or better than ones from other programs (quality), and if materials were used by other institutions (sustained use). Essentially, we did not identify if good, worthwhile, or important work continued, but rather if continuation happened at all.

In another evaluation of an international development program, sustainability was not the key emphasis of the evaluation, but was still considered as an integral part of the programming, thus of the impact evaluation. Sustainability in this instance is inherent not only in the continuation of a program or the implementing agency (which may or may not be relevant in a given region, once the needs of the people are addressed), but manifest in meeting the needs of people while considering the ability of future generations to meet their own need. In a nutshell, the program provided livestock to members of a larger group of families. The whole group of farmers received training in all aspects of animal care, in addition to learning about plant production, environmental maintenance, improvement of social issues such as equity, the importance of education for children of both genders, and means for developing micro credit systems within the group. However, the key sustainability element of the program is that offspring of the livestock provided to some of the group members would be passed on to remaining families within the group until the whole group received animals. The expected outcomes were improved economic, social, and environmental conditions of families, larger groups of individuals, and whole communities now and in the future. The evaluation yielded evidence of maintained benefits to recipient families which was commonly, but not always, extended to the other group members via the pass on of the offspring animals and the creation of micro credit systems. In some cases, the impacts were replicated within larger community which in any case received indirect benefits. As such, sustainability was not only viewed in terms of continuing
activities of the implementing agency, but of continuing and replicating impacts on (i) recipient families and respective groups, (ii) on the larger community, and (iii) potential other communities (Schröter, 2008a, pp. 218-219).

As illustrated, both examples present very different manifestations of sustainability, which need to be further explored. The first example illustrates program continuation, while the second addresses the larger concerns related to meeting human, social, economic, and environmental needs. Related back to the noun sustainability, two core perspectives of sustainability evaluation emerge: (1) evaluation of sustainability, that is, the capacity to continue something; and (2) evaluation for sustainability, that is, the maintenance or maximization of human, social, and economic positive impacts under consideration of environmental necessities. Thus, in its literal meaning, sustainability can be interpreted as both a process and an outcome. To address both issues requires evaluation (i.e., the determination of merit, worth, and significance) of sustainability (of evaluands) for sustainability (human survival on earth).

Sustainability and Sustainable Development

Historically, sustainability has been a concern as long as humankind has exists and intervened with the environment (Johnson, 2000; Glasby, 2002; Wheeler,

---

2 Detailed timelines on the history of sustainability and sustainable development are available from the International Institute for Sustainable Development (IISD), The Rochester Institute of Technology (RIT), and several other Web sites.
In the last fifty years, sustainability received increased attention worldwide via events such as the United Nations Educational, Scientific, and Cultural Organization (UNESCO) International Conference for Rational Use and Conservation of the Biosphere in Paris, France (1968); the UN Conference on the Human Environment in Stockholm, Sweden (1972); the formation of the World Commission on Environment and Development (WCED, 1983); the United Nations Conference on Environment and Development (UNCED) in Rio de Janeiro, Brazil (1992); the establishment of the Commission on Sustainable Development (CDS, 1993); the Johannesburg Summit in South Africa (2002); and the creation of the United Nation’s Millennium Development Goals (MDG); to name but a few. In the beginning of this movement, the major focus was on global environmental and economic issues, while later on, the connections to human and social dimensions of sustainability were increasingly emphasized.

The most widely known and used definition of sustainability stems from this sustainable development movement and was published in the Brundtland report, *Our Common Future* (1987). Thereafter, sustainable development was defined as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs” \(^4\) (WCED, 1987, p. 43).

---

\(^3\) Acronyms used throughout the dissertation can be found in Appendix A.

\(^4\) This definition is the most prominent in the sustainable development literature and stems from the so-called Brundtland Report, developed by an independent committee of twenty-two experts and headed by the Prime Minister of Norway, Gro Harlem Brundtland, to scrutinize global environment and development issues to the year 2000 and beyond. Critical problems, proposals for solving them, and increasing recognition are at the core of the report. The first part addresses most common concerns that
This definition covers issues related to development at all levels of developed and developing societies and emphasizes dynamics between social and economic development and environmental protection (c.f., Dresner, 2002; Bell & Morse, 2003; Gibson et al., 2005). The social, economic, and environmental aspects are commonly referred to as the pillars of sustainable development (United Nations, 2002). Many other definitions of sustainability have been formulated within the sustainable development movement and variations to the three pillar approach exist (see Chapter II). Some speak of capital models (e.g., Pezzey, 1992), others distinguish human sustainability from social sustainability (Goodland, 2002), and yet others add further dimensions such as politics, culture, empowerment, or education (e.g., Schubert & Störmer, 2007).

Definitions of sustainability emerge as different perspectives or philosophies about sustainable development. For example, Hardi (2007) stresses three macro perspectives on sustainability: (a) an ecological interpretation in which sustainability is defined as resilience or maintained dynamic capacity of a system to adapt to changes and disturbance; (b) a transition perspective, where sustainability is perceived in terms of continuous changes or social advances and in which sustainability is used synonymously with sustainable development (cf., Meadowcroft, 2007; Parris & Kates, 2003a); and (c) an economic interpretation that considers
substitution within different capitals. Various other definitions exist for different levels (global, regional, national, local) of program implementation (c.f., Störmer & Schubert, 2007), that is on mega, macro, and micro levels.

Macro discussions tend to emphasize global sustainability issues such as global warming, overpopulation, overconsumption, and ecological footprints among others (e.g., Meadows, Randers, & Meadows, 2004). These are widely addressed via global guidelines (e.g., Agenda 21) and national-level policies and programs (e.g., National Sustainable Development Strategies [NSDS], Poverty Reduction Strategies [PRS]) (Dalal-Clayton & Bass, 2006).

Micro perspectives focus specifically on local, project-level, or single-site interpretations. For example, sustainability of social and international development interventions are commonly defined in terms of continuing human, social, economic, and environmental benefits from a development intervention in a given area under set conditions (OECD, 1992, 1998; DANIDA, 2006). A definition by the United States Agency for International Development (USAID, 1988) suggested that a development program has been sustainable when it is able to deliver an appropriate level of benefits for an extended period of time after major financial, managerial and technical assistance from an external donor is terminated. The Organisation for Economic Cooperation and Development’s Development Assistance Committee (OECD DAC, 1992) defines sustainability as concern with measuring whether the benefits of an
activity are probable to continue after donor funding has been withdrawn, while considering environmental and financial sustainability of projects.

Common themes within international, sustainable development frameworks include holistic, long-term perspectives; systems thinking; accountability and shared responsibility; human, social, and economic development within the realm of environmental protection; equity; and participation, cooperation, and collaboration, among others (Aaker & Shumaker, 1996; Hardi & Zdan, 1997; Leiserowitz, Kates, & Parris, 2006; Redclift, 1987; Reid, 1995; Ukaga & Maser, 2004). Cracknell (2000) emphasized seven requirements for sustainability in international aid, including (i) the need for government support regardless of the level of implementation (i.e., grassroots versus national program); (ii) the need for managerial ability and effective institutions; (iii) the need to select appropriate technologies for a given country; (iv) the need for sociocultural compatibility, local participation, and gender awareness; (v) the need for environmental compatibility (regardless if the project appears to be environmentally neutral or not); (vi) the need for financial viability; and (vii) the need for robust project designs able to cope with the unexpected.

In contrast to discussions in the realm of sustainable development (SD), evaluation-specific journals (e.g., Evaluation and Program Planning or the American Journal of Evaluation) describe sustainability most frequently in terms of program continuation, community capacity, or the extension of outcomes (c.f., Shediac-Rizkallah & Bone, 1998; Reinke, 1999; Wandersman et al., 2000; Howard &
Howard, 2000; Johnson et al., 2004; Rog et al., 2004; Scheirer, 2005). Others stress that sustainability begins early in a program’s lifecycle (c.f., Mancini & Marek, 2004; Pluye et al., 2004, 2005; Bailey et al., 2003). In essence, sustainability in these cases is either viewed as continuation of processes or outputs. Prior to the completion of the initial funding cycle, sustainability is a procedural construct developed via leadership competence, collaboration and networking, understanding of the community, demonstrating program results, strategic funding, staff involvement and integration, and program responsiveness. After the initial funding cycle, sustainability represents continued output, characterized by (a) unremitting or repeated implementation of services, (b) maintenance of community capacity, or (c) institutionalization, routinization, or standardization. Additionally, socioeconomic, political, and environmental considerations, program duration, training, education, administrative structures, interpersonal relations, transparent communications, technology, and risk taking are considered important in some of these instances (e.g., Mancini & Marek, 2004). Alignment with consumers’ needs and demonstrated positive evaluative results are less frequently mentioned.

Sustainability Evaluation

Corresponding to the ubiquitous nature of the concept of sustainability, evaluations thereof encompass a great range of evaluands, including policies, systems, practices, institutions, communities, services, programs, projects, to name
but a few. Sustainability pertains to all evaluands that are intended to meet the needs of humanity now and in the future. As systematic determination of merit, worth, and significance, evaluation is a tool for informing decision making. The most widely known functions of evaluation are these:

1. Summative evaluation with the intent to inform decision making (by funders) about program continuation, termination, or replication (Scriven, 1991) by “determining the value and future of the program and model” (Patton, 2008, p. 140)

2. Formative evaluation with the intent to inform learning (Patton, 2008) and decision making about the potential for improving the evaluand (Scriven, 1991)

3. Ascriptive evaluation with the intent to generate knowledge (Patton, 1996; 2008; Chelimsky, 1997) that may or may not lead to decision making or program improvement (Coryn, 2007)

Most recently, Patton (2008) distinguishes three further purposes, including accountability, monitoring, and development. Characteristics of these purposes are:

4. Accountability with the intent to inform decision making of those responsible for managing, legislating, and funding a given evaluand. Resource-effectiveness, efficiency, and goal achievement are in the center of this purpose (p. 140)
5. Monitoring with the intent to inform internal decision making and accountability by continuously measuring performance on sets of indicators (i.e., performance measures)

6. Developmental evaluation is concerned with an evaluand’s ability to adapt in “complex, emergent, and dynamic conditions” (p. 141).

All six of these purposes have implications for sustainability evaluation, not all of which are explicitly discussed in the pertinent sustainability evaluation literature. Summative evaluation is concerned with whether a program should be sustained. Formative evaluations may ask how to improve a given evaluand to enable sustainability or better meet the requirement for sustainable development. Note, not all programs are intended to be sustained beyond the funding cycle. Instead, there are programs that are planned to be completed at some point. For these programs, sustainability is implied in the maintenance or continuation of outcomes with the purpose to thwart reverse effects on achievements. Ascriptive evaluation would look at contributors and inhibitors of sustainability and be concerned with questions derived in terms of evolutionary sustainability (Kraft & O’Neil, 2007). For example, at what point are we evaluating something different from what was initiated originally? Accountability-oriented evaluations would, for example, look at program fidelity, which may or may not be relevant for sustainability evaluation. Finally, developmental evaluation is concerned with adaptations of an evaluand in a given
context, which also relates to notions of evolutionary sustainability. Such adaptations are specifically relevant for the larger SD movement.

Understanding sustainability and potential questions asked in evaluating sustainability opens the door to the world of SD for both evaluators and programmers. The conceptualization of sustainability as introduced within the Sustainability Evaluation Checklist (SEC) developed within this dissertation intends to bridge the gap between mega, macro, and micro perspectives, and to focus on community-level development interventions and differentiate two approaches to evaluating sustainability:

1. Sustainability is a retrospective, outcome-oriented concept in which positive intended and unintended outcomes are continued beyond the removal of initial program resources and the immediate reach of the evaluand. In this regard, sustainability entails continuation, replication, and exportability of program activities and impacts.

2. Sustainability is a moral principle or global goal requiring any development intervention to produce lasting positive outcomes on human, social, economic, and environmental dimensions. From this perspective, sustainability is a process-oriented, formative concept that has to be considered prospectively throughout a program’s life cycle and beyond.
The checklist also emphasizes general considerations in evaluation which might be of relevance to some but not all users, depending on their training and experience in conducting evaluation.

**Problem Statement**

In sum, sustainability is a means to an end as much as it is worthwhile striving for as a moral principle. Evaluators of interventions in general and sustainable development efforts, specifically, should be aware of key concepts of sustainability and consider all aspects relevant to sustainability within their evaluations. At this time, sustainability evaluation too often is only perceived from one direction and assessed on one dimension. Either sustainability is evaluated in terms of the continuation of projects or part of projects beyond the initial funding cycle and evaluated accordingly. Or sustainability of long-term impacts is evaluated on a singular dimension—environmental, social, or economic. The concept of sustainability, however, is more complex. Interdependencies between dimensions exist and unidimensional evaluations fail to capture the whole picture.

This is similar for evaluations on differing levels (mega, macro, micro). Global efforts tend to be subject to top-down policy evaluation on the mega and macro levels, while program sustainability evaluations emphasize the micro level and too often are only concerned with program continuation and institutionalization beyond initial funding cycles. It is argued that SD requires both.
The SEC, developed and validated within the realm of this dissertation, intends to fill the void and discrepancies by providing a comprehensive approach to evaluating sustainability, because sustainable development requires both consideration of continued activities and outcomes beyond initial funding cycles and concern with long-term impact on human, social, economic, and environmental dimensions of sustainability. Specifically, the SEC intends to link levels and dimensions of sustainability. On the one hand, it aspires to be a generic framework for evaluation efforts in various contexts that gives attention to most, not just a few, aspects that should be considered in sustainability evaluation. On the other hand, the checklist is intended for use in stand-alone sustainability evaluations as well as in evaluations that consider sustainability as a subcomponent. The SEC is not intended to argue that anything has to be sustained, but rather to support decision making about whether sustainability is desirable or not and how the evaluand can be improved to maximize sustainable development potential. In a recent review on the state of sustainable development, it was clear that the quality of sustainability evaluation efforts has been mixed to date (UN AMR, 2008).

Aim and Scope of the Dissertation

This dissertation explores sustainability evaluation from multiple angles and proposes a tool intended for use in planning and designing project and program evaluations of sustainability (retrospectively) for sustainability (prospectively) within
development contexts. As a device that helps users to remember certain tasks involved in evaluation and that considers a wide array of criteria of importance to sustainability evaluation, the checklist aims to reduce errors of omission and increase evaluation usefulness. In addition to planning and designing sustainability evaluations, the SEC should also support the generation of ideas and discussion on key issues in sustainability evaluation, facilitate proposal-writing processes, and allow for comparing existing sustainability evaluations to determine whether all important aspects have addressed.

Built on Scriven’s (1995, 2005) and Fournier’s (1995, 2005) logic of evaluation, the SEC is intended as a guide for individuals who are involved in internal or external evaluations of development projects or programs with a special interest in sustainability. These users may include evaluators and researchers who provide evaluation services; program planners, funders, and managers with an interest in evaluation; program recipients/participants/users who conduct their own evaluations; and others who have an interest in evaluation.

Usually, evaluation checklists are developed, validated, and refined by their usage over long periods of time and have credibility because of the extensive experience, knowledge, and professional status of their originators (Stufflebeam, 2000a; Scriven, 2007). Within his Checklist Development Checklist (herein referred to as E-CDC), Stufflebeam (2000a) suggests a set of procedures consisting of twelve checkpoints to be considered when developing and validating checklists. These were
used as the larger frame for this dissertation, in addition to a framework put forth by Hales, Terblanche, Fowler, and Sibbald (2007).

Organization of the Dissertation

To set the frame for developing the SEC, Chapter II, *Sustainability Evaluation*, is concerned with focusing the checklist task by synthesizing the pertinent evaluation literature to clarify and define the content area *evaluation* under specific consideration of the aforementioned conceptualizations of *sustainability* (OECD DAC, 1992; Hardi & Zdan, 1997; Caspari, 2004; Stockmann, 2006; Mancini & Marek, 2004; Pluye & Potvin, 2004, Pluye, Potvin, Pelletier, Mannoni, 2005; Thierstein & Walser, 2007; Ridde, Pluye, Johnson-Lafleur, 2007). Specific attention will be paid to definitions, purposes, evaluands, components, dimensions, and the logic of evaluation (Scriven, 1995, 2005; Fournier 1995 & 2005) as used within the SEC.

In Chapter III, *SEC Development* procedures are discussed. This chapter begins by providing an overview of checklists, in general, by addressing three questions: (1) What are evaluation checklists? (2) Why develop them? (3) What other sustainability checklists are in existence? After answering these questions, this chapter describes the methodology of checklist development, compares two different models for developing checklists (Stufflebeam, 2000a; Hales et al, 2007), considers issues related to formatting checklists (Hales et al, 2007; Biechelmeyer, 2002; 2003),
and discusses procedures used for developing the SEC. The third section provides a summary of key concepts used in the checklist and elaborates on the checklist’s content. The draft SEC used in the validity study can be found in Appendix B.

Chapter IV, methodology, introduces the methodology for evaluating the SEC. After clarifying the purpose, sampling procedures, recruitment strategies, and mechanisms for protecting human subjects are discussed. The next part of this chapter includes the research procedures, wherein the data collection methods, instrumentation, and measures are illuminated and pretest and pilot test procedures and results are described. At the end of Chapter IV, the location of the data and data-analytic strategies are discussed.

Chapter V presents the results from the validation study. First, respondent characteristics are summarized. Second, the validity rating scales comprised of an accuracy scale and a utility scale are evaluated in terms of their reliability and validity. Special emphasis is given to the two validity subscales “accuracy” and “utility” as well as other checklist characteristics that were highlighted by the respondents. Third, attention shifts to the qualitative data gathered via interviews, the questionnaire, and e-mail feedback. Findings from cross-case, cross-item, and cross-section analysis are presented to illuminate strengths, weaknesses, and recommendations for improving the SEC. Both qualitative and quantitative findings are triangulated in a fourth section of this chapter, which concludes with a summary and implications for checklist improvement.
The final chapter (Chapter VI, Discussion and Conclusions) begins with a discussion of the revisions made to the SEC, including changes to the content, format, and structure. This section also highlights aspects that were not addressed in revising the SEC and provides rationales for why they were not addressed at this time. The revised SEC can be found in Appendix L. The second section of Chapter VI calls attention to the limitations of the dissertation and corresponding implications for future work. This section is followed by a discussion of the contributions of this dissertation. The dissertation ends with concluding remarks and future directions.

The evaluation community may benefit from this research by receiving a newly developed and validated tool that will potentially guide efforts for evaluating sustainability. Also, the study may yield important knowledge that can be used for validating evaluation checklists in the future. This dissertation is limited primarily because it was impossible to conduct field tests of the SEC. Funding limitations and time constraints did not allow for such investigation.
II. SUSTAINABILITY EVALUATION

Evaluation is a transdiscipline that studies and improves certain tools for other disciplines while often remaining autonomous with respect to its structure and research efforts (Scriven, 1991, 2003). At the heart of professional evaluation is the systematic, objective determination of merit, worth, and/or significance. Merit is the intrinsic goodness of the evaluand without specific consideration of cost. Worth pertains specifically to cost and is defined as the goodness of the evaluand under consideration of cost. Significance is the importance of the evaluand. The evaluand is that which is being evaluated (c.f., Scriven, 1991, 2006; Joint Committee, 1994; Davidson, 2005a). In serving this function, evaluation is both a ubiquitous and practical endeavor in which evaluators often function in multiple roles, using multiple methods from an array of disciplines.

To illustrate the transdisciplinary nature of evaluation, Scriven (1991, 2003, 2004) introduced the metaphor of the “country of the mind” (p. 13). Building on this metaphor, it can be argued that evaluators work in the “house” of evaluation (see Figure 1).

This house has a ground floor, a second floor, and an attic. The ground floor is the evaluator’s primary workplace in which all applied endeavors take place, such as

---

5 This metaphor is also expanded upon by Coryn and Hattie (2006) in their discussion of the transdisciplinary model specifically in terms of the evaluation of research.
methodological developments for specific evaluands. The evaluand (object under evaluation) is situated in applied fields such as sustainable development and characterized by one of the evaluation fields (e.g., program, product, policy, or systems evaluation, among others). In the second floor, a handful of evaluators generate disciplinary theory. The attic is reserved for metatheory, which holds those elements that seem implied to those working in the other floors of the house, but which sometimes need to be looked into to remain on track. Although evaluators work in the house of evaluation, usually they were born and live under the roofs of other disciplines.

![Figure 1. The House of Evaluation](image)

Other disciplines (economy, sociology, management, linguistics, environmental sciences, psychology, …)
This is specifically apparent in sustainable development, where evaluators are economists, political scientists, human rights advocates, and gender experts, among many others. When these evaluators come to work, they bring the wisdom of their house and their friends’ houses, but they also share the knowledge from work, so that all households utilize evaluation to some extent in their decision-making processes. In essence, the use of evaluation is so ubiquitous across applied fields in all disciplines that it is often forgotten as such.

This ubiquity is specifically relevant for sustainability. Like evaluation, issues related to sustainability pervade most areas of human endeavor. Organizations want to survive; program managers want to sustain processes and outcomes; politicians want to remain in power; environmentalists want to fortify resources; teachers want to sustain knowledge; and above all, humanity desires to survive on planet Earth. As such, sustainability is a relevant issue for most evaluands, not to mention for evaluation itself.6 Sustainability evaluations, thus, encompass a great range of evaluands in multidisciplinary contexts, including policies (e.g., Rao, 2000; Mudacumura, Mebratu, & Shamsul Haque, 2006; Strong & Hemphill, 2006); systems (e.g., Alperovitz, 1996); practices (e.g., Aaker & Shumaker, 1996); institutions (e.g., Martinuzzi, 2002), communities (e.g., Quablan, 2005); and services, programs, and projects (e.g., Howard & Howard, 2000), and so forth.

---

6 While sustainability of evaluation is not the topic of this dissertation, some efforts in this direction have been made (see Miller, Kobayashi, & Noble, 2006).
This chapter summarizes key concepts and concerns with regard to sustainability evaluation within the larger frame of the logic of evaluation. For the forthcoming *Sage Glossary of the Social and Behavioral Sciences* (Sullivan, Johnson, Mercado, & Terry, forthcoming), I (Schröter, in press) describe the logic of evaluation (based on Scriven, 1991, 1995, 2005; Fournier, 1995, 2005) as follows:

The logic of evaluation pertains to its definition, definitions of its major concepts, its nature and relationship to other disciplines, and the rules of inference that govern it. The *definition* of evaluation is based on three evaluative predicates: (i) merit (goodness without consideration of cost); (ii) worth (goodness under specific consideration of cost); and (iii) significance (importance or relevance). Although many definitions of evaluation exist, those are often too narrow (e.g., limiting evaluation to general social/behavioral science research) or too broad (e.g., taking anything that generates useful information as professional evaluation).

There are three overarching *purposes* of evaluation: (i) improvement (formative evaluation); (ii) accountability (summative evaluation); and (iii) knowledge generation (ascriptive evaluation). Evaluation objects are referred to as *evaluands*; evaluation subjects are called *evaluees*. Evaluation is practiced in varying disciplines and fields of study in the arts, humanities, social and behavioral sciences, engineering, and business, for example. *Branches* of evaluation include policy, program, personnel, performance, product, portfolio, proposal, intradisciplinary, and metaevaluation.

The distinguishing feature between evaluation and research is the centrality of *valuing* in evaluation. In contrast to the value-free doctrine commonly practiced in the social sciences, evaluators purposefully ascribe value. The working logic of evaluation consists of four major steps: (i) establishing criteria of merit (i.e., on what components or dimensions must the evaluand do well?); (ii) constructing standards (i.e., how well should the evaluand perform?); (iii) observing/measuring performance and comparing to standards (i.e., how well did the evaluand perform?); and (iv) synthesizing and integrating information into an evaluative conclusion (i.e., what is the relative or absolute merit or worth of the evaluand?).

*Fundamental operations* for determining absolute or relative merit, worth, or significance of an evaluand can be subdivided into (i) *primary operations* (grading, ranking, scoring, apportioning, and synthesis) and (ii) *secondary operations* (weighting—assigning importance levels, barring—setting minimum performance levels which cannot be compensated for by
better performance on any other dimension or component, and profiling—the graphical illustration of performance).

Professional evaluators strive for essential claims that are based in scientific evidence and require the establishment of factual and value premises. Factual premises comprise descriptive information, while value premises provide relevant and indisputable standards that are applied to the facts to establish conclusions about absolute or relative merit, worth, and significance of an evaluand. These describe the key function of evaluative inference, namely to move validly from factual and definitional premises to evaluative conclusions. Other value claims are open to dispute in evaluation, including personal preference (subjective claims that are not definitionally true and cannot be validated), market claims (that can be verified by common sense and law, and contextual claims (prima facie factual claims that hold true in a given context) (Schröter, forthcoming).

Key issues within this description will be further elaborated below with specific attention to sustainability.

Definitions

‘Contributions to sustainability’ is frequently presented as an official objective of environmental assessment . . . . But true sustainability assessments, carefully designed and intentionally influential, are still rare (Gibson, Hassan, Holtz, Tansey, & Whitelaw, 2005, pp. 2-3).

The SD literature points out that “evaluation is useful for decision makers particularly because it helps in understanding what SD means in operational terms” (Hardi, 2007, p. 28). Yet, “assessment” is the predominant term used for sustainability evaluation. While some debates about terminology suggest that assessment is different from evaluation in that it (a) emphasizes quantitative approaches and (b) would pertain to personnel evaluation specifically (see Scriven,
this is not the case in sustainability evaluation (e.g., Hardi & Zdan, 1997; Gibson et al., 2005), where assessment and evaluation are used interchangeably. First, sustainability evaluation generally understood as a type of program, policy, or project evaluations. Second, both quantitative and qualitative means for assessing sustainability exist. Thus, assessment and evaluation will be used synonymously herein.

Definitions of evaluation are generally varied. Hardi (2007) argues, for example, that evaluation is (a) an explanatory tool for SD; (b) a planning tool that informs programming practice; and (c) a performance measurement tool that assesses goals and objectives of SD, where SD itself is not clearly defined and requires interpretation, operationalization, and choice of method by the evaluator. This definition will not be used here, because evaluators generally agree that explanation, planning, and goal achievement may be purposes or side effects of evaluation, but do not comprise definitional properties (Scriven, 1991; Scriven, 1994a; Stufflebeam, 2007). Hardi also argues that evaluation processes are content-dependent and vary according to different definitions of SD. While this is true for the choice of method, it stands in contrast to the existence of a universal logic of evaluation that is applicable across contexts (Fournier, 1995).

Although evaluation has been growing as a professional discipline and practice for more than forty years, definitional ambiguity continues to exist. Table 1

---

7 Examples for qualitative “assessments” are presented by Ormala (1994) or Berger & Steurer (2006) for example.
lists a few definitions from sources generally concerned with sustainability evaluation.

### Table 1

**Definitions of Evaluation**

<table>
<thead>
<tr>
<th>Definitions of evaluation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. to determine as systematically and objectively as possible the relevance, efficiency,</td>
<td>UNEP, 2005, p. 55</td>
</tr>
<tr>
<td>effectiveness and impact of the organizations' activities in relation to their objectives</td>
<td></td>
</tr>
<tr>
<td>2. an assessment, as systematic and impartial as possible . . . It focuses on expected and</td>
<td>UNEP, 2007a</td>
</tr>
<tr>
<td>achieved accomplishments, examining the results chain, processes, contextual factors and</td>
<td></td>
</tr>
<tr>
<td>causality, in order to understand achievements or the lack thereof. It aims at determining</td>
<td></td>
</tr>
<tr>
<td>the relevance, impact, effectiveness, efficiency and sustainability of the interventions</td>
<td></td>
</tr>
<tr>
<td>and contributions of the organizations of the UN system.</td>
<td></td>
</tr>
<tr>
<td>3. to determine as systematically and objectively as possible the extent to which project/</td>
<td>UNEP, 2007b</td>
</tr>
<tr>
<td>programme needs and results have been or are being achieved, and to analyse the reasons</td>
<td></td>
</tr>
<tr>
<td>for any discrepancy</td>
<td></td>
</tr>
<tr>
<td>4. to assess, monitor, grade, or judge the intended outcomes, and/or unintended effects</td>
<td>Ukaga &amp; Maser, 2004, pp. xi-xii</td>
</tr>
<tr>
<td>of specific activities on a particular system . . . Evaluation of sustainable development</td>
<td></td>
</tr>
<tr>
<td>emphasizes the process of decision making based on the analysis of perceived risk.</td>
<td></td>
</tr>
<tr>
<td>5. refers to the process of determining the worth or significance of an activity, policy</td>
<td>World Bank IEG, 2007, p. 13</td>
</tr>
<tr>
<td>or program. An assessment, as systematic and objective as possible, of a planned, on-going,</td>
<td></td>
</tr>
<tr>
<td>or completed intervention.</td>
<td></td>
</tr>
<tr>
<td>6. (1) assesses the effectiveness of an ongoing program in achieving its objectives, (2)</td>
<td>United States Environmental Protection Agency, 2007</td>
</tr>
<tr>
<td>relies on the standards of project design to distinguish a program's effects from those</td>
<td></td>
</tr>
<tr>
<td>of other forces, and (3) aims at program improvement through a modification of current</td>
<td></td>
</tr>
<tr>
<td>operations</td>
<td></td>
</tr>
<tr>
<td>7. systematic and objective assessment of an on-going or completed project, program or</td>
<td>OECD, 2002, pp. 21-22</td>
</tr>
<tr>
<td>policy, its design, implementation and results. The aim is to determine the relevance</td>
<td></td>
</tr>
<tr>
<td>and fulfillment of objectives, development efficiency, effectiveness, impact and</td>
<td></td>
</tr>
<tr>
<td>sustainability. An evaluation should provide information that is credible and useful,</td>
<td></td>
</tr>
<tr>
<td>enabling the incorporation of lessons learned into the decision-making process of both</td>
<td></td>
</tr>
<tr>
<td>recipients and donors. Evaluation also refers to the process of determining the worth</td>
<td></td>
</tr>
<tr>
<td>or significance of an activity, policy or program. An assessment, as systematic and</td>
<td></td>
</tr>
<tr>
<td>objective as possible, of a planned, on-going, or completed development intervention.</td>
<td></td>
</tr>
</tbody>
</table>

As shown, the United Nations Environment Program (UNEP) alone lists three different definitions of evaluation. All three of these definitions agree that evaluation
should be systematic and objective. The first emphasizes both significance (i.e., relevance) and worth (i.e., efficiency), but seems simplistic in terms of merit, which might subsume effectiveness and impact, but should extend to other factors such as ethicality for example. The second definition only considers anticipated and accomplished success, which leaves doubt about the function of evaluation and raises the question of why unintended outcomes and impacts are not a concern. Similar to the first definition, importance, effectiveness, impact, and efficiency are named. Additionally, sustainability is perceived as a criterion for evaluation in the UN system. The third definition emphasizes programmatic needs rather than the needs of those the program intends to affect and considers whether objectives were achieved.

Within the discipline of evaluation, the most widely known and used definition is the systematic and objective determination of merit, worth, and/or significance or the product thereof (Davidson, 2005a, p. 240; Scriven, 2005, p. 255). Merit represents intrinsic quality without consideration of cost. In essence, is the evaluand sustained or sustainable? Does it have sustainable impacts? Worth considers quality in terms of cost specifically. Basically, is the evaluand sustainable at reasonable costs? Do costs and resources allow for the achievement of sustainable impacts? Significance is used synonymously to importance. Essentially, how important is the sustainability of the evaluand and its effects under consideration of
the needs of those to be served and the risk environment? Coryn (2007) argues that significance “is the overall conclusion when all relevant considerations have been synthesized” (p. 30). However, herein, significance is defined as an equal function to merit and worth. Gibson and colleagues (2005) state:

Significance is involved [in sustainability assessment] because at every step of the process there is too much to do. There are too many undertakings, too many possible alternatives, too many potentially relevant factors, too many conceivable effects, and too many options for mitigation, enhancement and adaptation. Not all of them can be considered, at least not thoroughly, and not all of the ones that are considered can be given much attention in the decisions on approval and implementation. Priorities must be set (Gibson et al., 2005, p. 166).

That is, the significance function is internal to the evaluation process and allows disentangling more important aspects of the evaluand from less important factors (e.g., via needs and values assessment). Thus, for sustainability evaluation as used here, the triumvirate values imply that evaluation is concerned with determining whether certain human actions (policies, programs, projects, etc.) have the requisite importance (i.e., significance) and quality (i.e., merit) to be continued or maintained within the constraints of available human, social, economic, and environmental resources (i.e., worth).  

---

9 Note that the Program Evaluation Standards (1994) do not explicitly refer to significance, but rather subsume it under merit. More detailed discussion on distinguishing characteristics of merit and worth are discussed by Stufflebeam and Shinkfield (2007, p. 10).

9 Note that importance is often associated with politics, but as described here, it relates to needs and risks.
Purposes

Purposes of evaluation are multifaceted and linked to uses and users of evaluation (Patton, 2008), directions in which evaluations are requested, and approaches that are chosen to conduct evaluation. However, Posavac and Carey (2003) state, “There is only one overall purpose for program evaluation activities: contributing to the provision of quality services to people in need” (p. 13). Within sustainability evaluation, this notion must be expanded to incorporate the environmental context.

The literature suggests at least six purposes of evaluation, including formative, summative, ascriptive, developmental, and accountability evaluation, as well as monitoring (see Chapter 1; Chelimsky, 1997; Coryn, 2007; Patton, 1996, 2008; Scriven, 1991). Each of these purposes is relevant to sustainability. The World Bank’s IEG (2007) names five additional purposes, including ethical, managerial, decisional, educational, and motivational reasons (see p. 16), which comprise subdimensions of the larger purpose categories.

Evaluation purposes are highly context-dependent. For sustainability assessments, the following purposes have been presented:

1. Improve decision making\textsuperscript{10} on all undertakings that may, individually or in combination, have a significant effect on progress toward sustainability

\textsuperscript{10} This represents the formative function.
2. Ensure comprehensive and integrated attention to all factors affecting long-term as well as immediate desirability and durability\textsuperscript{11}

3. Provide the core framework (structure, criteria, and process) for deliberation and decisions on significant undertakings\textsuperscript{12}

4. Encourage overall consistency and efficiency in decision making from policy to program design to post-approval project implementation and monitoring, while also favoring flexibility and decentralization by respecting uncertainty and context, working iteratively with the relevant stakeholders, and adapting to different ecosystems and communities, new understandings, and emerging challenges and opportunities\textsuperscript{13}

5. Encourage effective public engagement in the conception, planning, approval and implementation of undertakings that may have a significant effect on progress toward sustainability\textsuperscript{14}

6. Foster and facilitate creative innovation as well as transitions to more sustainable practices (Gibson et al., 2005, p. 148)

While it is arguable whether each of these represents purposes of evaluation rather than criteria for evaluation implementation, formative and developmental functions are evident as dominant themes.

\textsuperscript{11} This reflects criteria for adequate evaluation (i.e., be systematic and objective). See specifically, the Joint Committee on Standards for Educational Evaluation

\textsuperscript{12} See footnote 11.

\textsuperscript{13} See footnote 11.

\textsuperscript{14} See footnote 11.
Evaluands\textsuperscript{15}

In this dissertation, emphasis is given to projects and programs and, to a lesser degree, policies—or PPPs (Eggers, 2006).\textsuperscript{16}

\textit{Policies}

Policies comprise "standards, guidelines, or rules set up by an organization to regulate development decisions" (World Bank IEG, 2007, p. 18). As such, they are the most pervasive form of programming and linked to strategic planning. Owen (2006) distinguished three types of policies:

(1) Legislative policy which is set by a government bill or parliamentary act, enabling allocation of funds

(2) Large-scale policy, developed by a governmental department enabling system-wide interventions under one organizational umbrella

(3) Local policy, developed by local organizations to facilitate single-site implementation of an intervention. (pp. 23-26)

Most SD policies relate to the latter two. In conjunction with the Brundtland Report (WCED, 1987), Agenda 21 (UN, 1997) and the local Agenda 21\textsuperscript{17}, many

\textsuperscript{15} A historical overview of different evaluands is provided by Coryn (2007).

\textsuperscript{16} After discussing the evaluands overall, this dissertation will use "PPP" and "program" to refer to sustainability evaluands.

\textsuperscript{17} "Local Agenda 21 is a local-government-led, community-wide, and participatory effort to establish a comprehensive action strategy for environmental protection, economic prosperity and community well-being in the local jurisdiction or area. This requires the integration of planning and action across economic, social and environmental spheres. Key elements are full community participation,
national strategies (NSDS) and poverty reduction strategies (PRS) were developed which constitute policies as used herein (cf., Kopp, 2006; Moffatt, 1996; OECD, 2006; Thin, 2002). In contrast to strategic plans, these policies are informed via policy analysis and research, which provides the frame for the following characteristics of “contemporary” policies:

- developed based on evidence
- focused on change
- recognizes multiple policy interests (inclusive)
- mission- and-vision oriented (strategic)
- interested in causal linkages between implementation and outcomes
- links cross-departmental or cross-sectoral boundaries
- consideration of implementing agencies/constituencies (realistic)
- applicable in a range of contexts (flexible)
- focus on benefits for policy clients (outcome-oriented)

(Owen, 2006, p. 26)

As summarized by Scriven (1991), policy evaluation emerged in the 1960s and 1970s through “operations research, microeconomics, organizational theory, public administration, social psychology, and the increasing interest in the role of law in public policy” (p. 267) and shares a number of characteristics with program evaluation (e.g., logic and methodology). Policy evaluation is conducted either

---

assessment of current conditions, target setting for achieving specific goals, monitoring and reporting” (http://www.gdrc.org/uem/la21/la21.html).
retrospectively (ex post) or prospectively (ex ante), where prospective policy
evaluation frequently addresses the “now what?” types of evaluative questions, which
often require sophisticated micro-simulations, formal modeling techniques, or
scenario constructions.

Policy evaluation is one core branch of evaluation in which sustainability is of
major importance. As discussed by Strong and Hemphill (2006), a large set of
policies have been put forth in SD, including policies on biodiversity, climate change,
construction, energy, environment, planning, pollution, social issues, transport, urban
development, waste management, and water to name but a few. Influences on
sustainability policies include the following (Rao, 2000):

1. The precautionary, or “better safe than sorry” principle was the foundation
   for substantive international laws adopted nationally, regionally, and
   locally. It contends that “uncertainty should be interpreted toward a
   measure of safeguard” (p. 100) and is an integral part of Agenda 21.

2. The International Organization for Standardization (ISO) put forth the ISO
   14000 family to inform ecology, economy, and macroeconomics in terms
   of environmental management (see Tibor & Feldman, 1996).

3. Fiscal policies are affected by ecotaxes or green taxes.

4. Market and market-based instruments are developed to influence the
   efficiency of environmental and other markets.

5. Institutional reforms would be necessary for sustainable development, but
   changes in the “human enterprise” are perceived as sluggish.
The most significant priorities that have shaped SD policies include:

- social priorities such as intragenerational equity, gender equity, health, inclusion of youth and children, eradication of illiteracy and poverty
- environmental priorities such as maintaining ecosystem resilience, improvement of environmental public health, preservation of biodiversity
- economic priorities such as meeting basic human needs, poverty reduction, and economic growth.

These reflect three-pillar (e.g., Dresner, 2002; Bell & Morse, 2003; Gibson et al. 2005) or triple-bottom line (e.g., Henriques & Richardson, 2005; Savitz & Weber, 2006) approaches widely discussed in the literature.

**Programs and Projects**

Scriven (1991) defines programs as “the general effort that marshals staff and projects toward some (often poorly) defined and funded goals” (p. 285) and projects as “time-bounded efforts, often within a program” (p. 286). The World Bank defines a program as “an intervention comprising various activities or projects, which are intended to contribute to a common goal” and projects as “single intervention in one location or a single project located in several locations” (World Bank IEG, 2007, p. 18). Owen (2005) suggests that programs consist of two primary elements: (1) a documented plan and (2) action consistent with the plan. He further distinguishes programming on three levels (see Table 2).

---

18 For more detail and elaboration, see Rao (2000) and Munier (2005) and this dissertation’s discussion of dimensions.
The mega level is similar to policy and generally influences macro-level programming, which usually influences micro-level projects. However, projects can also be developed bottom-up or inside out—initiated by grassroots efforts that are not necessarily funded.

### Table 2
Programming Levels

<table>
<thead>
<tr>
<th>Program planning levels</th>
<th>Who</th>
<th>What</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mega</strong></td>
<td>Governmental or corporate level</td>
<td>Planning of overall social, economic, and environmental impacts</td>
<td>Agenda 21</td>
</tr>
<tr>
<td><strong>Macro</strong></td>
<td>Divisions, department, regions, groups within an organization</td>
<td>Multi-site</td>
<td>NSDS &amp; PRS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agricultural programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Social programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Educational programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Health programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Others</td>
</tr>
<tr>
<td><strong>Micro</strong></td>
<td>Work units, individuals</td>
<td>Single site</td>
<td>Implementation on a community level</td>
</tr>
</tbody>
</table>

Within evaluation theory, programs receive the most attention, thus program evaluation has the most developed principles and practices (Owen, 2005; Coryn, 2007). Most sustainability evaluation relates to programs, as discussed in the subsequent section.

---

19 Madaus and Stufflebeam (2000) place the origins of professional program evaluation in 19th century Great Britain where attempts to reform education, law, hospitals, orphanages, and public health were...
Sustainability as Evaluand

When considering the literature pertaining to sustainability and sustainable development, it becomes evident that publications in evaluation-specific journals are more concerned with micro-level sustainability, while the SD literature emphasizes the macro and mega levels. Articles published in evaluation-specific journals tend to define programs (organizations or projects) loosely as systems under which action may be taken toward a goal. Within this literature, sustainability is most frequently described in terms of (a) program continuation beyond termination of initial funds, (b) community capacity, or (c) the extension of outcomes (c.f., Howard & Howard, 2000; Johnson et al., 2004; Reinke, 1999; Scheirer 2005; Wandersman et al., 2000). The most commonly used framework for operationalizing sustainability is from Shediac-Rizkallah and Bone (1998), who summarized six variants for defining sustainability that can be subsumed under the following categories:21

---

21 Note that program continuation is not desirable for the sake of it. As Shediac-Rizkallah and Bone (1998, p. 88) state: "...there are circumstances under which the discontinuation of a program is appropriate." In that sense, consideration of the larger SD context is paramount.
1. Maintaining services, thus continuing outcomes
2. Continuation beyond removal of resources from the initial funder
3. A process inherent in organizational change by means of institutionalization, routinization, and standardization
4. Extraorganizational capacity building

Later publications in the evaluation literature pertain to one or more of these categories, although none of these capture sustainability in terms of sustainable development specifically. Instead, the program itself is the focus of what is to be sustained. For example, Howard and Howard (2000) used the Self-Determination/Reliance Model (SDR) as a means to monitor movement on dimensions of accountability, decision making, information, knowledge and skills, and resource mobilization toward sustainability by community groups. Johnson and colleagues (2004) define sustainability as “the process of ensuring an adaptive ... system and a sustainable innovation that can be integrated into ongoing operations to benefit diverse stakeholders” (p. 137) and suggest a theory-driven sustainable planning model for infrastructure capacity building and sustainable innovation confirmation. Rog and her colleagues (2004) investigated sustainability of collaborations and stress factors that predict sustainability of these entities in terms of resources (monetary and nonmonetary) and degrees of operational, consistent, and growing continuation of extraorganizational and interorganizational efforts. Scheirer (2005) investigated empirical sustainability studies and defined sustainability as that which happens after initial external funding for a program ends, including (i)
continuation of program activities, (ii) maintained community capacity, and, to a lesser extent, (iii) continuation of benefits for new clients as investigated. Pfuye and colleagues (2004, 2005) stress that planning for sustainability beyond the initial funding cycle or beyond immediate effects of program activities starts at the onset of programs (c.f., Mancini & Marek, 2004; Pluye et al., 2004, 2005) and includes a focus on process-oriented strategies (c.f., Bailey et al., 2003). None of these scholars explicitly consider the larger sustainable development movement, thus are in danger of “dwarf[ing] the environmental dimension” of sustainability (Jänicke, 2007, p. 42).

In his Key Evaluation Checklist (KEC), Scriven (2006) refers to sustainability as a facet of generalizability. As such, sustainable parts of an evaluand can be generalized internally by substitution of resources or resource renewal, or externally by replicating, exporting, or expanding elements of an evaluand. Stufflebeam’s (2007b) Context, Input, Process, and Product (CIPP) Evaluation Checklist includes a whole subevaluation on sustainability, which was added in the fifth iteration of the CIPP Model in 2002. In the checklist, sustainability evaluation is defined as the assessment of “the extent to which a program’s contributions are successfully institutionalized and continue over time” (p. 9).

In sum, although, sustainability is discussed in the evaluation literature, it does usually not take the larger SD concerns into account. Neither has it been adequately defined for evaluation, but rather operationalized for unique contexts. This lack of definition is particularly evident in the absence of entries on sustainability, sustainability evaluation, or sustainable development in seminal texts such as

The capacity to exist after external support is completed or withdrawn; temporal durability through autonomy [1]. Often taken to be equivalent to “resilience to risk” but not quite the same, since the latter is more general and requires a general “bulletproof” quality with respect to unexpected disasters including natural disasters [2], whereas the former refers mainly to continuation after the termination of external *funding*. However, it should be noted that, in practice, we often expect sustainability to include the capacity for *political* and perhaps also environmental autonomy as well as fiscal autonomy [3]. In recent years, sustainability has become a major item of concern in the design and evaluation of international development projects, partly because it was for long ignored with the consequent recurrent failure of development projects to produce any lasting benefits for the recipient nations or communities. Since planning for sustainability is now important, its handmaiden, evaluation for sustainability, should now be a requirement for competent evaluation in the international sphere. One important component of such evaluation is attention to the necessary infrastructure, which is often a contextual rather than an internal requirement for success, but must be specified in the evaluation [4].

One [1] represents the notion of evaluation of sustainability, which is most widely discussed in American evaluation literature; [2] and [3] introduce the notion of evaluation for sustainability; and [4] represents a relevant criterion for both evaluation of and for sustainability.

---

22 The Evaluation Center is currently in the process of developing an Evaluation Wiki.
23 Personal communication on 3/12/2006.
24 The bracketed numbers are included as points of reference.
Within the realm of SD, Hardi (2007) argues that the object of evaluation is dependent upon the definition of SD, which he clusters in four broad categories that have different implications for sustainability (see Table 3).

Table 3

SD Categories and Characteristics (Based on Hardi, 2007) and Example Definitions

<table>
<thead>
<tr>
<th>Categories</th>
<th>Characteristics/Assumptions</th>
<th>Some example definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD as lifestyle</td>
<td>• Humans and their activities cause degradation in a world of finite resources</td>
<td>• Sustainability is the maintenance and continuation long-term health improvement of human and ecological systems (Wheeler, 2004, p. 24)</td>
</tr>
<tr>
<td></td>
<td>• Global political processes aim to change human lifestyles</td>
<td>• Continuing community capacity, social institutions, and/or social practice (Meadowcroft, 2007)</td>
</tr>
<tr>
<td></td>
<td>• Global scientific processes aim to increase understanding of limitations set by finite resources</td>
<td></td>
</tr>
<tr>
<td>SD as process/transition</td>
<td>• Emphasis on intrinsic characteristics of SD processes</td>
<td>• “sustainable development is a process and not a destination” (Brandon &amp; Lombardi, 2005, p. 26).</td>
</tr>
<tr>
<td></td>
<td>• Multi-scale, multi-domain, multi-temporal structural societal changes</td>
<td>• “a social construct, referring to the long-term evolution of a hugely complex system—the human population and economy embedded within the ecosystems and biogeochemical flows of the planet” (Meadows, 1998, p. 7)</td>
</tr>
<tr>
<td>SD as economic issue</td>
<td>• Substitution within different forms of capital</td>
<td>• Economic sustainability as substitution within different capitals, cost of replacement via substitution (shadow) prices that do not vary with changes in valuation by the market, inclusive wealth (Hardi, 2007)</td>
</tr>
<tr>
<td></td>
<td>• Problems in terms of putting monetary figures on all human, social, and environmental capitals</td>
<td></td>
</tr>
</tbody>
</table>
Table 3—continued

<table>
<thead>
<tr>
<th>Categories</th>
<th>Characteristics/Assumptions</th>
<th>Some example definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD as developmental issue</td>
<td>• Specific targets and timeframes</td>
<td>• Sustainability is the assessment of the potential continuation of positive outcomes of an activity once donor funding has been removed. In a postscript, the definition emphasizes environmental and financial sustainability of projects (OECD, 2007b)</td>
</tr>
<tr>
<td></td>
<td>• Maintenance of functioning, efficiency, and accessibility of resources, services, infrastructures, income generation, and community cohesiveness (based on CIDA, 2002 in Hardi, 2007)</td>
<td></td>
</tr>
</tbody>
</table>

Using the lifestyle perspective; the Brundtland definition (WCED, 1997) would not accommodate evaluation while the thermodynamic systems view would, specifically under consideration of substitution. Explicitly, needs satisfaction as addressed by the Brundtland definition would be limited by resource constraints that are set by the boundaries of the natural environment in relation to available technologies to allow for alternate resource use. This is where discussions of weak and strong sustainability are of importance. Weak sustainability exists when (specifically natural) resources can be replaced by means of man-made capitals, while strong sustainability implies that substitution of resources is not possible at this time (e.g., Gibson et al., 2005; Hardi, 2007).

A transitions definition would best suit evaluation. From a transitional perspective, sustainability is viewed as a moral principal and process that should guide human action (Hardi, 2007; Parris & Kates, 2003a; Pezzey, 1992). Specifically, sustainability is manifest in a community’s (or program’s or policy’s) control over

25 See also Jones, Young, and Stanley (2004)
natural, human, manufactured, social, and cultural capital in pursuit of economic
security and social justice (democracy), “while maintaining the integrity of the
ecological systems upon which all life and production depend” (Viederman, 1995, p. 46). That is, sustainability is a multi-scale, multi-domain, and multi-temporal
construct, implying continuous change (Hardi, 2007).

Change is not a straight, linear process as suggested by most development
models; rather it is interconnected and cyclical. Both negative and positive
results are possible, indeed probable, though one can neither predict all
outcomes, nor attribute all change to ones’ interventions (Aaker &
Shumaker, 1996, p. 3).

The economic view favors cost analysis in which nonmonetary costs and risks
can be accounted for via substitution. At the core of the discussion is maintenance or
growth of capitals (Goodland, 2002). Hardi (2007) argues that neo-classical
economists recently achieved a breakthrough in term of analyzing SD. While
economists previously were unable to consistently cost out natural, social, and human
capitals, neo-classical economic reforms illuminated means for placing monetary
values on “key capital stocks in nature, human welfare and human knowledge” by
introducing shadow pricing (p. 22).

In the case of development programs, evaluation would focus on institutional
or community-level interventions that could be assessed by comparing outcomes with
clearly set goals on multiple levels and in a set timeframe (Hardi, 2007). The
international development perspective is exemplified by Aaker and Shumaker (1996):

Sustainable development involves strategies which ensure long term
ecological, economic and social stability. These might include technologies
and social structures from the outside, but also often point to the return to
older, indigenous, more natural and culturally based methods. Often it is a blending of the two.

Sustainable development also involves the nurturing of appropriate values as well as individual and group empowerment. As an ecological perspective emerged, it has become imperative to mobilize communities to avoid destructive exploitation of natural resources and work toward more just, equitable and sustainable livelihoods for all (p. 2).

However, linkages between the international development perspective and the transitional perspective exist (Aaker & Shumaker, 1996; Patton, 1994) and are represented in evolutionary sustainability models (Kraft & O’Neill, 2007).

In sum, this discussion reflects the ubiquitous and ambiguous nature of sustainability: “sustainable development is the complex of activities that can be expected to improve the human condition in such a manner that the improvement can be maintained” (Munro, 1995). As stated by Gibson and colleagues (2005), “Sustainability stands as a critique; it is a challenge to prevailing assumptions, institutions and practices” (p. 38).

As a moral principle as well as a necessity, it is paramount that sustainability and SD are considered within program evaluation. Furthermore, it is not enough to conceptualize sustainability simply as continuation or existence of program impacts. Instead, sustainability is also the ability to adapt to changing conditions (Patton, 1994). To facilitate the integration of both concepts, evaluation of sustainability for sustainability is promoted herein. Evaluation of sustainability is defined in terms of PPP maintenance and continuation, while evaluation for sustainability is defined in terms of a moral principle, SD, to strive toward.
Evaluation OF sustainability
The determination of the merit, worth, and significance of efforts to continue a given evaluand (usually beyond the removal of initial resources)

Evaluation FOR sustainability
The determination of the merit, worth, and significance in maintaining, replicating, and exporting a given evaluand’s positive (un)intended outcomes and impacts under specific consideration of global sustainability issues.

Prior to illuminating the logic of sustainability evaluation, a brief excurse will be taken to elucidate various components of sustainability evaluands, as well as sustainability evaluation consumers and impactees.

Components of Sustainability Evaluands

Given that sustainability is usually connected to some type of human activity, such as that represented by PPPs, the following components can be derived: inputs, activities, outputs, outcomes, and impacts. Examples are illustrated in Table 4.

Inputs include material, nonmaterial, monetary, and nonmonetary resources of an evaluation object (Davidson, 2005a; Scriven, 1991). A key assumption related to inputs in terms of sustainability is their renewability (see strong versus weak sustainability).

Activities are actions that are assumed by an evaluation object to achieve specific goals (Frechtling, 2007). These actions vary tremendously in sustainability evaluation, depending on the sector and associated key sustainability concerns. This
dissertation is not intended to limit investigations to one specific sector, due to the varied nature of evaluands in sustainability evaluation and existing cross linkages between evaluands and sectors. As such, the framework developed here “incorporate[s] adaptive flexibility and respect[s] the specifics of context” (Gibson et al., 2005, p. 11).

Table 4

Components of Sustainability Evaluands

<table>
<thead>
<tr>
<th>Components</th>
<th>Example: International programming (highly simplified)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>Donations, values, knowledge, skills, etc.</td>
</tr>
<tr>
<td>Activities</td>
<td>Provision of products (e.g., goods or livestock) and services (e.g., education, information, etc.)</td>
</tr>
<tr>
<td>Outputs</td>
<td>Numbers of individuals/families who received products and/or services</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Changes in lifestyle due to activities and outputs (e.g., family/community nutrition, income, etc.)</td>
</tr>
<tr>
<td>Impacts</td>
<td>Long-term changes in community, sub-national levels, nation</td>
</tr>
</tbody>
</table>

*Outputs* are tangible, immediate results that comprise evidence of the implementation of an activity or service (Frechtling, 2007; Mathison, 2005). These may include numbers of individuals who received services, numbers of activities implemented, and so on. Outputs generally do not indicate whether a given intervention is any good, but rather signify the extent to which activities or services were implemented or provided.

*Outcomes* usually comprise intended changes, but should also include unintended change occurring as a consequence of the evaluand’s activities (Davidson,
Unintended changes are of specific relevance, because they inform an evaluands merit or worth.

*Impacts* include intended, unintended, anticipated, and unanticipated effects on targeted and nontargeted populations; usually referring to long-term effects and outcomes (Davidson, 2005a; Frechtling, 2007). Impacts are highly linked to sustainability. In essence, the notion “evaluation for sustainability” emphasizes impact evaluation specifically.

**Stakeholders and Impactees**

Several terms have been put forth to describe consumers of and potential participants in evaluation. These include clients, audiences, stakeholders, and consumers (see Table 5).

Stakeholders are individuals, groups of people, or organizations that have a vested interest (i.e., stake) in evaluation processes and findings. They can range from donors to recipients and bystanders of development interventions (Cracknell, 2000; Davidson, 2005a; Greene, 2005; Patton, 2008). Greene (2005) distinguishes four specific groups of stakeholders: (a) those that have decision power (e.g., donors, policymakers, funders, and advisory boards); (b) those with direct responsibility for a program (e.g., program planners, administrators, managers, and staff) (c) program recipients and downstream impactees’ and (d) those who could have been served (lost funding opportunity), though other interested parties might exist and constitute
additional stakeholder groups. Figure 2 displays relationships between different stakeholder groups under consideration of space, time, and nature of effect.

Table 5
Definitions of Users, Clients, Audiences, Stakeholders, and Consumers

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>“Those specific stakeholders selected to work with the evaluator throughout the evaluation to focus the evaluation, participate in making design and methods decisions, and interpret results to assure that the evaluation is useful, meaningful, relevant, and credible. Primary intended users represent key and diverse stakeholder constituencies and have responsibility for transmitting evaluation findings to those constituencies for use” (Patton 2008, p. 72).</td>
</tr>
<tr>
<td>Clients</td>
<td>Those who have “commissioned an evaluation and to whom the evaluator has legal responsibility; not the employer of whoever hires or, often the instigator of the evaluation . . . immediate consumers (i.e., recipients) of the evaluation” (Scriven, 1991, p. 82).</td>
</tr>
<tr>
<td>Audiences</td>
<td>“Those who will or should read or hear of the evaluation, either during or at the end of the evaluation process, including many who are and many who are not being evaluates” (Scriven, 1991, p. 62).</td>
</tr>
</tbody>
</table>
| Stakeholders       | “One who has substantial ego, credibility, power, futures or other capitals invested in the program, and thus can be held to be to some degree at risk with it . . . Recipients are only stakeholders in an indirect sense and normally dealt with separately . . . taxpayers” (Scriven 1991, p. 334-335)  
“Those people who are personally involved with the program, who derive some of their income from the program, whose future career might be affected by the quality of the program, or who are clients or potential recipients of the program’s services” (Posavac & Carey, 2003, p. 30) |
| Consumers          | “Anyone affected by a program or product, directly or indirectly, intentionally or unintentionally” (Scriven, 1991, p. 98).                                                                                                                                                              |

The degree to which stakeholders and impactees are involved in evaluation depends on the context. According to Patton (2008), “stakeholder involvement has become accepted practice in the profession” (p. 79). In sustainability evaluation, participation is paramount and has been established as one of the overarching values
to yield best use of results (e.g., Cracknell, 2000; Gibson et al., 2005; Stockmann, 2006; Thierstein & Walser, 2007).

**Figure 2. Relationships Between Different Stakeholder Groups**

In essence, “broad participation” constitutes one of *The Bellagio Principles for Assessment* (Principle 8):

Assessment of progress toward sustainable development should:

---

26 *The Bellagio Principles for Assessment* were developed by an international group of measurement experts that was called by the WCED in November 1996 to examine and synthesized assessment efforts to that date. The ten principles are intended to guide the assessment of sustainable development in communities, nongovernmental organizations, national governments, and international institutions. Four facets of the assessment process are distinguished, including (a) the establishment of a vision and goals (principle 1), (b) contents (principles 2 to 5), (c) the assessment process (principles 6 to 8), and (d) capacity building (principles 9 and 10), (Hardi & Zdan, 1997; Hardi, 2007).
• obtain broad representation of key grass-roots, professional, technical and social groups, including youth, women, and indigenous people—to ensure recognition of diverse and changing values

• ensure the participation of decision-makers to secure a firm link to adopted policies and resulting action (Hardi & Zdan, 1997, p. 3)

Stufflebeam and Shinkfield (2007) also emphasize the importance of integrating the stakeholders in planning the evaluation as well as during the evaluation process to assure that information needs and expectations for the evaluation are met. Fitzpatrick, Sanders, and Worthen (2004) even present a checklist for identifying key stakeholders and other audiences in relation to functions these groups could serve during the evaluation. The functions include (a) creating policy, (b) making operational decisions, (c) providing input, (d) reacting to findings, and (e) “for interest only” (p. 202). Greene (2005) states that involvement is more than providing input: “Stakeholders who are involved in an evaluation process contribute to important decisions regarding evaluation planning, implementation, and use” (p. 397). There would be at least two functions of stakeholder involvement: (a) to enhance usefulness of the evaluation result, process, or both and (b) to advance values related to equity, empowerment, and social change within the evaluation context (p. 397). Patton (2008) discusses stakeholder involvement in terms of power and degrees of interest. As illustrated in Figure 3, the power and interest of different stakeholder groups affect how they might impede or facilitate an evaluation.
People who set the context (e.g., politicians, advocates, dissidents, protestors, oppressors)

Could pose problems to evaluation use

Evaluators should spark their interest to thwart risks

Primary intended users (e.g., funders, program managers, administrators, leaders)

Potential clients and high-impact users who may influence evaluation use, use it themselves, and/or disseminate findings.

Stakeholders and Impactees

General public stakeholders, potential indirect, downstream impactees, bystanders

Transparency to these potential audiences could prevent from threats and maximize potentials if these audiences move in either direction

Individuals who may be affected by findings (e.g., recipients, consumers, internal stakeholders)

Involvement in evaluation can increase diversity and build evaluation capacity

**Figure 3.** Power, Interest, and Stakeholders Involvement (inspired by and adapted from Patton’s [2008, p. 80] Power Versus Interest Grid)

**Dimensions of Sustainability**

Dimensions are those aspects of evaluands that permeate throughout different components. In sustainability evaluation, these include contextual influences, such as human culture, beliefs, and perceptions; social organization; the nature of economy and investment patterns; and environmental assets and challenges (local or regional), (Goodland, 2002; Hardi, 2007; Hawkes, 2001; Pezzey, 1992).

The maintenance or continuation of something worthwhile to humankind is most widely discussed in terms of social, economic, and environmental dimensions
(Bell & Morse, 2003; Dresner, 2002; Gibson et al., 2005). This tripartite value set is commonly understood as the pillars of SD (United Nations, 2002), although variations exist. Some argue that education, culture, and empowerment should be added as pillars in their own right, while others substitute the social pillar with equity (Wheeler, 2004).

Pezzey (1992) frames his interdisciplinary discussion about sustainability in terms of capitals, which are defined as “any economically useful stock” (p. 322) and distinguishes between (a) natural capitals (e.g., forest, groundwater), (b) physical capitals (i.e., human-made resources such as buildings, machines), (c) human capitals, and (d) intellectual capitals. Both human and intellectual capitals are manifested in knowledge and skills, and their combination with physical capitals comprises human-made capitals.

Goodland (2002) defines *human sustainability* in terms of maintaining or improving human capitals such as health, education, knowledge, and leadership and pays specific attention to basic human needs (such as food, shelter, health, etc.). Social sustainability builds on the human dimension and is defined in terms of maintaining or improving social capital (e.g., culture, language, shared rules, laws, etc.). Attention is paid to social needs (e.g., ways of organization, governance, and human interaction).

*Social sustainability* is affected by interdependencies between development and social norms (Hardi, 2007). Thus, sustainability of interventions is largely dependent on the degree to which activities match social norms without overstepping
a community's tolerance for change. Grounded in religion, tradition, and custom, not all social norms are reflected in law, but are evident in assumptions about equity and ethics and presented by individual and/or group behavior and attitudes. Aspects of social norms are manifest in language, education, interpersonal relations, existing hierarchies, and levels of tolerance, among other things that are not primarily economically motivated. Because social norms are ubiquitous, not clearly observable, and ever-changing, the determination and evaluation of social limits is complicated. Development that occurs too quickly can lead to increased uncertainty and instability within a community, which in turn can affect sustainability of beliefs, codes of conduct, and usefulness of norms (Munro, 1995). As a social and political process (Trzyna, 1995), development practitioners must consider which human and social aspects of a community should be sustained and which should be changed and what implications changes may have. In essence, social sustainability implies “psychological adaptation” (Pezzey, 1992, p. 324).

Goodland (2002) describes economic sustainability as keeping capital intact, with attention to economic needs, infrastructure, distribution of wealth, control over resources, and overconsumption. The economic perspective on sustainability is built on assumptions of continuous growth, growing wealth of some members of a society at the cost of increased poverty of others (distribution and equity), market economy, and competition (Hardi, 2007). Ecological economics is specifically concerned with reducing environmental damage and increased use of renewable resources to sustain reserves, which requires that additional costs incurred through environmentally and
socially sound practices are reflected in pricing, as they support economic and ecological sustainability (Munro 1995). In essence, “economic sustainability is constrained by anything that upsets a viable balance between benefits and costs” (Munro, 1995).

Some argue that economic growth is impossible under the auspices of sustainable development (Daly, 1997), while others counter that economic growth does not necessitate increased material and energy use, but rather implies increases in the value of outputs which would be possible at decreased material and energy use due to innovation (Munier, 2005; Pezzey, 1992).

*Environmental sustainability* is defined in relation to protecting natural capitals (e.g., water, land, air, minerals, etc.) and concerned with ecological needs (related to pollution, climate change, waste management, and green energy, for example), (Goodland, 2002). It usually refers to “physical and ecological limits to sustainability” (Pezzey, 1992, p. 324). Key challenges to achieving environmental sustainability include overconsumption and overpopulation, as illustrated in the following remarks by Pezzey (1992):

*Industrialism* cannot be statically sustainable, since it constantly depletes available reserves of environmental resources, especially non-renewable ones. Only the dynamic sustainability of a successful treadmill is possible: new reserves *must* be discovered, and new tools (machinery, buildings, vehicles, etc.) and knowledge *must* be accumulated, to avoid decline . . . . [However] technical innovation has been encouraged for its own sake, as well as to relieve resource constraints. (p. 329)

World population, having taken 10,000 years to grow from about 5 million to 1 billion in 1825, doubled in just the next 100 years, and is currently doubling in less than 50 years . . . . (p. 329)
Both overconsumption and overpopulation have detrimental effects on the environment, including the rapid depletion of renewable natural resources, such as forests and fish; known reserves of nonrenewable energy and minerals; and nonrenewable stocks of genetic diversity and soil (Pezzey, 1992, pp. 330-331). Additionally, pollution, increased inequalities between rich and poor countries, and the accelerated speed at which economies and societies change are named as key problems for environmental sustainability (Pezzey, 1992). Development under the auspices of environmental sustainability refers to activities that do not cause environmental degradation and maximize the utilization of renewable resources (Larsen, 1999).

A fifth dimension, presented by Hardi (2007) is ecological sustainability, defined as “resilience of the system: its maintained dynamic capacity to respond adaptively to changes and disturbance” (Hardi, 2007, p. 19). Under an ecological worldview, sustainability is perceived as a core value and goal that incorporates “recent developments in physics, ecology, and psychology along with core elements of many of the world’s great spiritual traditions . . . . it is a code word for other values” (Wheeler, 2004, p. 31) such as those located in environmental, economic, and social (equity and ethics) value domains (Munro, 1995).

One of the key concerns of the ecological perspective is the limitations of the ecosystem within which “vital life-supporting processes can take place” (Munro, 1995). As such, humans depend on air to breathe, fresh water to drink, food to eat, and raw materials with which to make clothing, construct shelters, cook, and heat.
Moreover, humans depend on earth’s regeneration mechanisms. While some cultures in developing areas are still able to live within the limits of their local ecosystem, more “developed” cultures consume and rely on resources from all over the world (Hawkes, 2001). Similarly, disruptions in the ecosystem can have far-reaching impacts; examples include acid rain, which can affect trees thousands of miles away, and the diminishing ozone layers (Munro, 1995).

The ecological perspective considers strong and weak sustainability, that is, the extent to which substitution of resources is possible within a system. Ideal sustainability occurs “when all consumption and absorption of ensuing waste occurs in the place where consumption directly occurs” (Qablan, 2005, pp. 15-16), thus can be achieved in a place-based context where the place establishes the ecological boundaries for possible production and consumption. The distinction provides implications for levels of importance in evaluation (i.e., weighting). That is, if resources cannot be substituted by current available technologies, their preservation appears more paramount in most cases.

Furthermore, sustainability evaluation requires consideration of temporal and spatial dimensions (Beck/ODI, 2006; Hardi, 2007). On a temporal dimension, PPPs may address short-term or long-term needs. While sustainability is usually related to long-term goals, humanitarian action, for example, explicitly addresses immediate needs for purposes of recovery or relief. Within this short-term context, sustainability has been reconceptualized as “connectedness,” defined as “the need to ensure that activities of a short-term emergency nature are carried out in a context that takes
longer-term and interconnected problems into account" (Beck/ODI, 2006, p. 27).

Adapted from the sustainability criterion promoted in the OECD DAC (1992), the connectedness criterion is related to considerations of relative expenditure, partnerships, and local capacity building (Beck, 2006; ODI, 2006).

The fourth Bellagio Principle takes the long-term SD perspective:

Assessment of progress toward SD should:

- adopt a time horizon long enough to capture both human and ecosystem time scales thus responding to needs of future generations as well as those current to short term decision-making
- define the space of study large enough to include not only local but also long distance impacts on people and ecosystems
- build on historic and current conditions to anticipate future conditions—where we want to go, where we could go (Hardi & Zdan, 1997, p. 2)

While spatial boundaries are most commonly set within the realm of human action on Earth27 (e.g., on global, regional, subregional, national, subnational, and local levels), time boundaries are more elusive (Hardi, 2007). Hardi points out that timeframes are implicit in political processes and cycles. For example, NSDS are framed within legislative cycles and political strategies, and local plans correspond to local politics. An underlying assumption is of course political power. Struhkamp (2007), for example, points out that efforts regarding SD in the United States "have been dismantled" (pp. 308-309) as a result of changes in the government from President Bill Clinton to President George W. Bush.

27 As humankind has gone beyond the explicit limits of the Earth, some disciplines require consideration of sustainability beyond this boundary (e.g., space exploration).
Intergenerational and intragenerational equity are two further dimensions widely discussed in term of SD (see for example Bellagio, Principle 3). For intergenerational equity, Gibson and colleagues (2005) maintain the following:

Favor present options and actions that are more likely to preserve or enhance the opportunities and capabilities of future generations to live sustainably (p. 103).

In contrast, intragenerational equity means ensuring 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, etc.) between the rich and the poor (Gibson et al., 2005, p. 101).

Both concepts relate to issues of social justice (Thin, 2002) where intragenerational equity is considered in terms of gaps between rich and poor and associated production and consumption patterns, health, employment, knowledge, and community security (Gibson et al., 2005). Gender equity and disparities between different age groups are not named by Gibson and colleagues, but comprise key issues within the concept. Intergenerational equity is more problematic, because justice is hard to determine prospectively. In essence, researchers ask what is fair for future generations. Specifically, the notion of intergenerational equity is criticized on the grounds of time being asymmetrical, thus implying that the so-called future

---

28 The third principle states that “assessment of progress toward sustainable development should: consider equity and disparity within the current population and between present and future generations, dealing with such concerns as resource use, over-consumption and poverty, human rights, and access to services, as appropriate; consider the ecological conditions on which life depends; and consider economic development and other, non-market activities that contribute to human/social well-being.”
generation does not have any decision-making power at this time (see Pezzey, 1992, p. 331).

Somewhat related is the concept and SD dimension self-reliance (see Galtung, 1980a, 1980b, 1980c; Preiswerk, 1980a, 1980b), which implies “more food for the neediest, better health for more people, more satisfaction and self-fulfillment in life, more confidence in oneself and in one’s community, and a more effective defence against exploitative forces” (Preiswerk, 1980a, p. 15). The rationale for the theories of self-reliance were based on the assumptions (Collins & Moore Lappé, 1980; Galtung, 1980b) that (1) priorities would change towards production for the basic needs of those most in need, (2) mass participation is ensured, (3) local factors would be better utilized, (4) creativity would be stimulated, (5) there would be more compatibility with local conditions, (6) more diversity of development, (7) less alienation, (8) ecological balance would be more easily attained, (9) important positive externalities would be internalized or given to neighbors at the same level, and (10) solidarity with others would be increased (see also Guiterrez, 2005)—to name but a few. However, self-reliance would also have negative effects such as (1) remaining inequalities at reduced inequity, (2) exploitation on all levels (local, national, regional), (3) decrease in organic ties, and (4) reduced mobility.

One specific development model in which sustainability is tied to the notion of self-reliance is Heifer International’s Cornerstones-based planning and management model (Aaker & Shumaker, 1996). The fourth cornerstone, sustainability and self-reliance, states,
HPI [Heifer Project International] funds projects for a limited time. The project groups must plan to eventually support themselves by member donations or fund raising activities, especially through productive activities. HPI has found that self-reliance is most easily achieved when a group has varied activities and finds support from several sources. (Aaker & Shumaker, 1996, p. 20)

Within the Heifer concept, self-reliance as a form of sustainability is expressed via diversification of sufficient monetary and nonmonetary operating support, numbers and/or types of self-initiated activities, increased numbers of individuals and families who benefit from the initiatives, and succession planning under specific consideration of decreasing the need for external assistance (see Aaker & Shumaker, 1996, p. 26).

The notion of self-reliance was widely discussed in the early 1980s and is said to be as old as humanity itself, especially as it related to food. Collins & Moore Lappe (1980) argue that food dependency, one of the key issues in SD, began with imperialism. Food self-reliance is dependent upon seven factors: (1) reallocation of control over agricultural resources to local levels; (2) mass self-initiative, instead of top-down (i.e., governmental) directives; (3) independence of trade; (4) bringing back together agriculture and nutrition; (5) revamping to an end, not a means, (6) revamping industry to serve agriculture, and (7) coordinated social planning.

Galtung (1980a) specifically considered self-reliance in terms of technology and voiced critiques that are still relevant to today's SD movements (almost thirty years later). In terms of basic human (material and nonmaterial) needs (economic dimension), he argued that technology would only satisfy those of the world elite
(intragenerational equity). On the social dimensions, structures compatible with meeting basic human nonmaterial needs would be lacking. Furthermore, he called for preserving ecological balances on the environmental dimensions. Both, the social and environmental dimensions would be necessary conditions for meeting basic human needs. Galtung furthermore discusses alpha (vertical) and beta (horizontal) structures of self-reliance, where alpha represents unequal exchange, vertical division of labor, dependency, fragmentation, marginalization, and segmentation. Beta structures, in contrast, comprise equal exchange, horizontal division of labor, autonomy, solidarity, participation, and integration. Arguments for one structure would result in arguments against the other. An adequate mix of both structures would allow for self-reliance. This mix would be to embed beta communities in an alpha structure, making the second as horizontal as possible. Table 6 lists suggestions for an alpha/beta mix of technologies, based on Galtung (1980, pp. 230-231).

<table>
<thead>
<tr>
<th><strong>Food</strong></th>
<th><strong>Modified alpha technologies</strong></th>
<th><strong>(Re)created beta technologies</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Reduce trade in food</td>
<td>• Try to restore the old system that food is grown within the horizon—local autarky</td>
</tr>
<tr>
<td></td>
<td>• Drop cash practices</td>
<td>• Patterns of local food preservation and storage</td>
</tr>
<tr>
<td></td>
<td>• Curtail international agribusiness</td>
<td>• Collective ground that can be used for food production</td>
</tr>
</tbody>
</table>
Table 6—continued

<table>
<thead>
<tr>
<th></th>
<th><strong>Modified alpha technologies</strong></th>
<th><strong>(Re)created beta technologies</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clothes</td>
<td>• Curtail international textile business</td>
<td>• Try to restore patterns of local handicraft for better quality and lasting clothes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Establish symbiosis with food production</td>
</tr>
<tr>
<td>Shelter</td>
<td>• Curtail international housing business</td>
<td>• Try to restore local house building pattern of local materials</td>
</tr>
<tr>
<td></td>
<td>• Make houses and habitats less center-periphery dominated</td>
<td>• Collective ground that can be used for housing</td>
</tr>
<tr>
<td></td>
<td>• Transfer more work to homes</td>
<td></td>
</tr>
<tr>
<td>Medical care</td>
<td>• Better distribution of centers for health care, rural clinics</td>
<td>• Emphasis on the beta community as positive health care participation, less distinction between healthy and ill</td>
</tr>
<tr>
<td></td>
<td>• Sanitation standards, control of drugs, control of epidemics</td>
<td></td>
</tr>
<tr>
<td>Schooling</td>
<td>• Better distribution of centers for schooling</td>
<td>• Emphasis on the beta community as source of education, integrating school, work, leisure</td>
</tr>
<tr>
<td>Transportation/communication</td>
<td>• Less centralized, two-way patterns between beta units</td>
<td>• Try to restore patterns of walking, talking (e.g., bicycles, wall posters); cars banned inside units; use of cable TV, local papers</td>
</tr>
<tr>
<td></td>
<td>• Collective means of transport, cars for long distance only</td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>• Better distribution of centers for large-scale energy production</td>
<td>• Solar, wind, biogas networks</td>
</tr>
<tr>
<td>Defense</td>
<td>• Democratized armies</td>
<td>• Local defense pattern, military and nonmilitary</td>
</tr>
<tr>
<td></td>
<td>• Better distribution of commanding positions</td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td>• Recycling</td>
<td>• Local control and more dependence on smaller economic cycles</td>
</tr>
<tr>
<td></td>
<td>• Cleaning-up technology</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>• Social transparency through citizen participation and reporting</td>
<td>• Small-size units comprehensible to anybody</td>
</tr>
</tbody>
</table>

As a result of the works presented in evaluation-specific journals, further dimensions of sustainability can be derived. Mancini and Marek (2004), for example, suggested seven dimensions pertaining to sustainability:
(i) leadership competence (strategic planning, evaluation, securing funding, supporting staff, etc.)

(ii) effective collaboration (inclusion of relevant stakeholders, communicating responsibilities clearly, sharing vision, etc.)

(iii) understanding the community (knowing community needs, understanding the context, culture, and politics, including community members, etc.)

(iv) demonstrating program results (formative evaluation to improve programs and secure future funding)

(v) strategic funding (planning ahead)

(vi) staff involvement and integration (commitment, qualification, development, etc.)

(vii) program responsiveness (flexibility to arising needs).

Each component includes several subcomponents that reflect core program characteristics that must be considered at different stages in program lifecycles (i.e., critical activities, immediate, intermediate, and long-term outcomes that may influence project sustainability and sustainment). These components are also inherent in several other studies focusing on sustainability. The evaluation-specific articles most frequently list factors such as resource stabilization through diverse funding streams and wide distribution of nonmonetary resources; community participation (networks and collaborations); alignment with strategic plans, visions, and missions; and leadership (a champion). Fullan (2005) described leadership at all levels as the lever for yielding sustainability. Also commonly referred to are monitoring and evaluation, financing, effectiveness, worthwhile outcomes, organizational capacity,
commitment and ownership, and knowledge and skills. Less frequently suggested aspects are socioeconomic and political considerations, program duration, training and education, administrative structures and linkages, adaptability, integration capacity, interpersonal relations, transparent communications, technology, cost effectiveness, transparency of benefits, risk taking, alignment with consumers’ needs, and negotiation. On the other hand, many of the less frequently addressed factors are identified as critical in the SD literature and summarized in Figure 4 (c.f., Gibson et al., 2005).

*Figure 4. Sustainability Requirements as Decision Criteria (based on Gibson et al., 2005, pp.116-118)*
Logic of Evaluative Reasoning

Regardless of how a given sustainability evaluand is operationalized, at the heart of evaluation is reasoning: "Ultimately evaluators reason their way to a concluding statement or series of evaluative conclusions as to the merit or worth of a program, product or person" (Fournier, 1995, p. 15).

From a pragmatic viewpoint, this means that evaluators work from factual and definitional evidence to valid (beyond reasonable doubt) evaluative conclusions (Scriven, 1991). This general logic of evaluation consists of four fairly straightforward steps that are illustrated in Figure 5 and discussed in the following sections.

Figure 5. General Logic of Evaluation (based on Fournier, 1995, p. 16)
In contrast to the general logic, the working logic of evaluation is

the variation in detail in which the general logic is followed when conducting an evaluation. In other words what varies across approaches is what or how criteria are identified, what or how constructs are constructed, how performance is measured, and how data are synthesized (Fournier, 1995, p. 18).

This working logic is highly diverse in different contexts (e.g., sectors, areas of application, or field) of sustainability evaluation. It is not emphasized here, because the checklist to be developed is intended for application in highly diverse settings and fields. The following section discusses the general logic of evaluation in relation to sustainability evaluation.

**Values and Criteria**

Values in evaluation provide the basis on which sustainability (or other evaluands or components and dimensions thereof) can be judged as good or bad (merit), worthwhile or worthless (worth), and vital or trivial (significance). They facilitate the movement from factual or descriptive data to evaluative conclusions. Valuing is the key distinguishing characteristic between evaluation and research.²⁹

²⁹ Evaluation theorist generally agree that evaluation differs from research in at least seven points: (1) purpose, (2) types of conclusions, (3) the role of valuing, (4) clients, (5) generalizability of results, (6) standards on which products from each process are judged, and the (7) preparation of professionals (Fitzpatrick, Sanders, & Worthen, 2004, pp. 6-7; Patton, 2008, pp. 40-41). Cronbach and Suppes (1969 as cited in Patton, 2008, p. 41) described the difference in terms of orientation, that is, while research is oriented toward conclusions, evaluation is oriented toward decisions.
Generally, four types of value claims are differentiated: personal preference, market, contextual, and essential value claims (Scriven, 1991, 2005).\textsuperscript{30} Personal-preference types of value claims are neither definitionally true nor can they be validated. For example, an individual may prefer blue cars over red cars or SUVs over motorcycles. They are in the eye of the beholder and do not relate to descriptions and explanations based on facts, hence they normally have no place in professional evaluation (c.f., Davidson, 2005a; Schwandt, 2005). They are, in part, the reason for the myth that values have no place in science.\textsuperscript{31}

Market value claims are those that can be verified or are recognized by the law and common sense, thus they are much less subjective than the personal values (Coryn, 2007). Such claims can often be found in terms of economic sustainability and are reflected in laws, policies, or taxes (French, 2005). For example, gas prices or environmental taxes reflect market values that do not hold true across nations, yet people accept these regulations in general.

Contextually evaluative claims are implicitly evaluative, but carry precise meaning in a specific context (Scriven, 1991). That is, these prima facie claims are characterized by their changing meaning in context. For example, the goals presented

\textsuperscript{30}Trzyna (2001) provides a definition for the SD context specifically, where values are “core beliefs or desires that guide or motivate attitudes and actions. Some values, such as the importance persons attach to honesty, fairness, and loyalty, are ethical in nature because they are concerned with the notion of moral duty—they reflect attitudes about what is right, good, or proper, rather than what is pleasurable, useful, or desirable.” This definition does not provide clarity about evaluative values.

\textsuperscript{31}For more detailed discussions on the topic of evaluation and the value-free doctrine, see Coryn (2007).
in the UN Millennium Development Goals (MDG\textsuperscript{32}) are contextually evaluative claims of special importance to nonindustrialized nations. However, what each MDG comprises in a specific country may vary. Another example for sustainability relates to the notion of overpopulation and overconsumption (Daly, 1997; Meadows, Randers, & Meadows, 2004) both of which have different meaning in differing contexts. In essence, some nations (usually developing nations) have extremely high birth rates and extremely low consumption rates, while others (usually industrialized nations) have extremely low birthrates, but are characterized by extreme overconsumption.

Most central to evaluation are essential value claims that require both factual and value premises, where factual premises are descriptive facts and value premises constitute the standards by which to determine the level or degree of merit, worth, or significance of the evaluand. Contextual and essential value claims are those embraced by objectivist evaluator as they are based on rationally defensible arguments that are descriptive and evaluative in nature (Davidson, 2005a; Schwandt, 2005; Scriven, 2005).

Criteria in evaluation are “aspects, qualities, or dimensions that distinguish a more meriteous or valuable evaluand from one that is less meriteous or valuable” (Davidson, 2005b, p. 91) or, in Scriven’s (2007) terms, primary indicators that are

\textsuperscript{32} There are eight MDGs: (1) eradicate extreme poverty and hunger, (2) achieve universal primary education, (3) promote gender equality and empower women, (4) reduce child mortality, (5) improve maternal health, (6) combat HIV/AIDS, malaria, and other diseases, (7) ensure environmental sustainability, and (8) develop a global partnership for development.
part of the concept of a “good X,” that are definitionally connected with the evaluand. These criteria can be derived from the sustainability components and dimensions discussed above.

Additional sources for identifying values and criteria are presented in Scriven’s KEC and in Stufflebeam (2001c) Values and Criteria Checklist as well as by Leiserowitz and colleagues (2006). Of special relevance for sustainability evaluation are needs of impactees (see Figure 2). They are critical sources of values and facts because the creation, existence, and sustainability of all evaluands are justified by their impact on end-users (Davidson, 2005a). There is general agreement that needs are context-dependent (c.f. Altschuld & Kramer 2005, Davidson 2005, McKillip 1998, Reviere et al., 1996, Scriven & Roth, 1978; Wiggins, 1998). They are also conditional and may evolve into wants or desires when a level of adequacy is reached (Scriven & Roth, 1978). The critical point at which needs must be met establish “the minimum standard for acceptability” (Scriven, 2006), or a “bar.” Needs are also distinguished across three dimensions (Davidson, 2005a; Scriven, 2006): (1) functional (a.k.a. performance) versus instrumental (a.k.a. treatment) needs (i.e., the actual or potential problem and the proposed solution); (2) conscious versus

---

33 Credible and valid criteria checklists are characterized by their criterial status; completeness, discreteness/independence, commensurability, clarity, conciseness, and conformability. In contrast, secondary indicators are factors, variables, or observations that are empirically connected with a criterion variable (Scriven, 2007). Unlike criteria, indicators are frequently unstable in their validity and can be “easily manipulated” (Scriven, 1991, p. 194). Bell and Morse (2003) suggest the following characteristics of indicators: Specific: must clearly relate to outcomes; Measurable: implies quantitative nature; Usable: practical; Sensitive: must readily change as circumstances change; Available: data must be collectable; and Cost-effective.
unconscious needs (those that can be identified by those in need versus those that cannot); and (3) met versus unmet needs (those that are addressed at the point of the assessment, but which must not be compromised, and those that are not addressed). If a functional need for sustainability can be identified, underlying instrumental means for meeting that need are implemented (i.e., programs or policies). Dependent on the specific context (i.e., applied field as well as location and time), the need for sustainability can be met or unmet and conscious or unconscious to those who have the need and to humankind in general (i.e., risks and vulnerabilities). In this regard, Altschuld and Witkin (1995, 2000) differentiate three levels of needs: (1) the primary direct consumer (e.g., those directly receiving services); (2) the provider group (e.g., midstream consumers); and (3) the system level (e.g., stakeholders). In addition, to the actual target group, further impactees are affected, sometimes indirectly or unintentionally (c.f., Davidson 2005, Scriven 1991) which is of special importance in sustainability evaluation.

Criteria of merit from standard usage and contextual meaning are another source for criteria and normally drawn from the evaluand. These relate to many of the sustainability dimensions described above and are exemplified by spatial and time dimensions, the risk of making irreversible changes, and consideration of sustaining indigenous cultures while, at the same time, improving general conditions and preserving the environment, and risks of repeating failures of the Western World. Unfortunately, it is argued that until now northern (or western) traditions of science have largely failed to value indigenous and experiential knowledge which may inform
the development of criteria from standard usage and contextual meaning for sustainability evaluation specifically (Viederman, 1995).

*Logical values* are formally true or valid requirements, in accordance with principles of reasoning on the basis of inference and demonstration. Examples of these can be drawn from the issues of overconsumption and overpopulation as presented in the *Limits of Growth* (Meadows et al., 2004).

*Legal values* are legislated or mandated requirements (e.g., environmental laws and regulations). These include for example the Atomic Energy Act, the Clean Water Act, and the Energy Policy Act to name but a few34 (see also French, 2005).

With regard to *ethical values*, Trzyna (2001) argues, "Only those value systems which subordinate non-ethical values to ethical ones are ethical." As a code of conduct based on moral duties and obligations, ethics deals with the capability to differentiate good from evil, right from wrong, and propriety from impropriety. As such, it is permeating all dimensions of consideration for sustainability evaluation. An example of the fundamental nature of ethics is presented by Daly (1991) who remarked: "the intrinsic value of other species, their own capacity to enjoy life, is not admitted at all in economics, and their instrumental value as providers of ecological life support services to humans is only dimly perceived" (p. 236). In essence, within sustainability evaluation, ethics is an omnipresent consideration to be made and

34 United States Environmental Laws and Regulations are available at: http://www.epa.gov/lawsregs/laws/index.html
affects all dimensions of merit, worth, and significance of sustainability and for sustainability.

Another source for deriving criteria are *political values*. Scriven (1991) suggests that it is often important to consider whether “the political value offsets the substantial inequality that it involves” (p. 207), implying an ethical dimension. In essence and as alluded to earlier, a simple change in political leadership may affect whether and what programming is implemented or not (cf., Struhkamp, 2007). Moreover, politics exist on all levels and internal and external to organizations. Wheeler (2004) argues that values are commonly shaped and propagated by societal institutions, including “social, political, cultural, and economic structures and traditions . . . [represented by] . . . systems of law, courts, and government; corporations, advertising, and media; and a large number of informal rules and codes of behavior.” (pp. 32-33).

*Resource economy* is another major source for values in sustainability evaluation. In essence, renewable resources (e.g., forest or fish) must be considered in terms of regeneration rates; nonrenewable resources (e.g., fossil fuels or fossil groundwater) must be considered in terms of substitutes and their potential for regeneration; and pollutants must not extent the level of recycling, absorption, or minimal harm (Meadows et al., 2004).

Values related to *risk or risk aversion* are another key considerations in sustainability evaluation and include for example risks of increased energy consumption, although improvements in energy efficiency are being made, risks in
blocking industrial innovation, and risks of causing irreversible changes to name but a few (OECD, 2001).

Other sources of values put forth include historical/traditional/cultural standards, scientific criteria and standards, technological values, marketability, personal and organizational goals/desires, fidelity, sublegal values, professional standards, and expert judgment (Scriven, 2006).

Values in evaluation should be consciously determined. First and foremost, these values are based in need. What really needs to be sustained and what is desirable but not critical? The severity and breadth of the need to be addressed by a given evaluand determines the level of importance for sustaining mechanisms or resources. Risk and vulnerability assessments determine the extent to which outcomes and impact might be detrimental if needs are not successful addressed (Davidson 2005; Gallopin, 2006).

35 These are critical for sustainability evaluation and relate to time/space and are context dependent.
36 Knowledge generated in terms of sustainability and SD.
37 Refers to potential users' access to the product or program or the plan for getting them used, in essence accessibility to knowledge and technology for sustainability.
38 These are assumed to be less important than needs; they often lack ethical or logical justification.
39 Compliance with plans for implementation. Fidelity is less important in cases where evaluands have to adapt to their environment (transitional and evolutionary perspectives on sustainability).
40 Normally not covered by legal or ethical values, in reference to important, often idiosyncratic, legislative preferences as opposed to those which are mandated. Could be plans created in relation to Local Agenda 21.
41 Values established by authority, custom, or general consent; most often as determined by a profession or professional association
42 Those of subject matter experts normally refer to the preferred standards of “experienced practitioners” (Picciotto, 2005, p. 31). However, expert judgment often takes the form of connoisseurship and is also subject to the fallacy of irrelevant expertise.
Within their framework, Gibson and colleagues (2005) describe the following conventional criteria for evaluating the significance of effects in sustainability assessments. These criteria are not comprehensive, but provide some general guidance:

- Stability and resilience
- Value and equity
- Unintended negative impacts and uncertainties
- Complexities and cross-linkages
- Permanence and irreversibility
- Intensity, magnitude, scale, extent, duration, frequency, and cumulation
- Levels of resource and energy consumption and waste production
- Existing standards, rules, laws, and regulations
- Human health and ecological risks
- Rareness, scarcity, and uniqueness of resources
- Level of public controversy

Standards

In contrast to the general value sources discussed above, standards comprise specific values that are applied and by which performance is assessed to determine the extent to which it is meriteous, valuable, or significant (Coryn, 2007). In

---

43 These are not stylistic standards, which are predicated on voluntary compliance (Picciotto, 2005). For example, the Joint Committee's standards for program and personnel evaluation are those by which evaluators should—voluntarily versus mandatorily or obligatorily—adhere (1988, 1994).
sustainable development contexts, these standards may comprise benchmarks or targets (e.g., UN MDG, 2007), which are commonly set for programs or organizations. While these benchmarks or targets are usually available and set in relation to organizational goals, evaluators may develop additional systems (e.g., rubrics or grading schemes) for judging the relative or absolute merit, worth, or significance of the evaluand. Fournier (1995) suggests illuminating warrants, backings, and conditions of exception. Warrants can legitimate evidence-based inferences by appeal to some authority. Backings support the warrant by appealing to a more general authority. Conditions of exceptions represent the limitations (i.e., circumstances when the warrant may not hold).

**Measure Performance and Compare with Standards**

It is not the primary interest of this dissertation to illuminate the general research practices, including data collection and analysis, used in sustainability evaluation. However, common techniques, in addition to the usual social science research methods (e.g., observations, focus groups, interviews, questionnaires, case studies), include tools specifically collected for purposes of sustainability assessment on the mega and macro levels (Arbter, 2007; Dalal-Clayton & Bass, 2006; Weaver, Rotmans, Tumpey, Haxeltime, & Jordan, 2007). For example, Dalal-Clayton and Bass (2006) reviewed the monitoring mechanisms for NSDS, specifically the use of national peer reviews, internal reviews, external auditing, parliamentary reviews, budgetary reviews, indicator-based and quantitative assessments, as well as public,
local, and international monitoring practices. Weaver and colleagues published are
large of sustainability assessment tools on the Web. Their site, Sustainability A-Test,
includes physical assessment tools (e.g., ecological footprint and lifecycle
assessment), monetary assessment tools (e.g., cost-benefits and cost-effectiveness
analysis, environmental accounting), modeling tools (i.e., integrated models, socio-
economic models, bio-physical models), scenario analysis tools, multi-criteria tools
(e.g., weighted sum, dominance method), vulnerability assessment tools (e.g.,
mapping techniques and indicator-based assessments), and participatory tools (e.g.,
electronic focus groups, interactive backcasting). The most widely discussed tools in
the sustainability literature are sustainability indices (Bossel, 1999; United Nations
Commission, 2001; Parris & Kates, 2003a, 2003b) and modeling approaches
(Moffatt, 1996; Moffatt, Hanley, Wilson, 2001).

Of general interest for the checklist to be developed are to broad categories of
evaluation procedures, that is, holistic and analytic evaluation (Davidson, 2005a).
Holistic evaluation, though implicitly analytical, attempts to determine the merit or
worth of an evaluand without consideration of the separate components or dimensions
of merit. In contrast, analytic types of evaluation explicitly differentiate dimensions
and components or dissect program theory for analytic purposes. In component
evaluation, each part of the evaluation object is evaluated separately. The resulting
subevaluations are then integrated into overall conclusions about the evaluand (see
Davidson, 2005a). Dimensional evaluation is concerned with determining merit,

44 See: http://ivm5.ivm.vu.nl/sat/
worth, and significance of facets that permeate the whole evaluand (see dimensions of sustainability discussed earlier in this chapter). Theory-driven evaluation essentially attempts to explain the causal linkages between linked components (see discussion on components) of an evaluand and "how and why" an evaluand "achieves a result" (Chen, 2005, p. 415).

Of specific relevance to sustainability evaluation is also the differentiation between prospective and retrospective evaluation, which explicitly involves the time dimension (see Figure 6).

*Figure 6. Life Cycles and Prospective/Retrospective Evaluation*
Prospective or, in European terms, ex ante evaluation is described as “somewhat similar to evaluability assessments” (World Bank IEG, 2007, p. 2007). Prospective evaluations usually are initiated to facilitate program planning or to inform program implementation formatively. In the sustainability context, prospective evaluations are conducted to build scenarios and predictive models to inform how efforts can be adjusted to maximize the greater good. That is, ex ante evaluation is conducted to predict likely future effects and inform policy development. Forecasting and forecasted returns on invested money are key issues in this type of evaluation.

In contrast to prospective approaches are retrospective or ex-post evaluation approaches. UNDP (2002) defines ex post evaluations as “a type of summative evaluation of an intervention usually conducted two years or more after it has been completed. Its purpose is to study how well the intervention (program or project) served its aims and to draw conclusions for similar interventions in the future” (p. 100-101).

Synthesis

Once all data are collected and analyzed, results have to be integrated in sets of ratings of performance on multiple dimensions or components to yield evaluative conclusions about merit, worth, and significance (Davidson, 2005). While synthesis occurs at several steps throughout the evaluation process (e.g., synthesizing facts with values, synthesizing multiple subdimensions on one criterion, etc.), some evaluations require an ultimate synthesis into one final judgment. If this is not the case, integrated
findings can be presented on criteria dimensions in form of profiles. Fundamental evaluative operations are necessary during the synthesis step and require advance consideration of the type of amalgamation sought. For example, grading and ranking operations lead to unidimensional or absolute conclusions, while profiling leads to multidimensional conclusions that illustrate grades and comparative performance on criteria.

**Summary**

While there are many efforts for evaluating *sustainability*, the general logic has not yet been explicated in the sustainability literature. Most efforts emphasize general research procedures that are related to, but do not exemplify, the general and working logic of evaluation. Thus, the logic of evaluation is an integral part in the checklist to be developed within the realm of this dissertation.

The next chapter illustrates the checklist development process by discussing (1) the background for checklist development, (2) the development methodology, and (3) key aspects of the sustainability evaluation checklist.
III. SUSTAINABILITY EVALUATION CHECKLIST DEVELOPMENT

The previous chapter summarized the key literature that informed the development of the SEC. This chapter begins by providing an overview of checklists, in general, by addressing three questions: (1) What are evaluation checklists? (2) Why develop them? (3) What other sustainability checklists are in existence? In the second section, this chapter describes the methodology of checklist development, compares two different models for developing checklists, considers issues related to formatting checklists, and discusses procedures used for developing the SEC. The third section provides a summary of key concepts used in the checklist and elaborates on the checklist's content. Finally, a summary intended to guide the reader into the next chapter is provided. Figure 7 provides a chapter overview.

![Figure 7. Chapter III Overview](image-url)
Background for Checklist Development

This section lays the foundation for the SEC’s development and asks the following questions: (1) What are evaluation checklists? (2) Why develop evaluation checklists? (3) What other sustainability checklists are in existence?

What are Evaluation Checklists?

In general, checklists are inventories consisting of “factors, properties, aspects, components, criteria, tasks, or dimensions” (Scriven, 2007, p. 1) that guide users in completing a given task by checking off items or referring to them during task completion (Stufflebeam, 2001a). In other words:

A checklist is an organized tool that outlines criteria of consideration for a particular process. It functions as a support resource by delineating and categorizing items as a list—a format that simplifies conceptualization and recall of information (Hales et al., 2007, p. 22).

Generally, experts agree about the mnemonic function of checklists. This means that checklists are perceived as practical tools that assist users in completing an assignment without forgetting key aspects or overemphasizing less important elements. In addition to including important items to consider in carrying out a particular task, checklists have other functions and features across disciplines (Stufflebeam, 2001a; Scriven, 2007; Hales et al., 2007). For example, they facilitate standardization. If a given checklist is used within an organization or among a group of professionals, the processes that it supports can be implemented consistently.
Checklist use can also increase objectivity among a group of users by assuring consideration of all points listed within the checklist without being sidetracked by personal assumptions about the task. Finally, checklist use can enhance the potential for replicating evaluations (Hales et al., 2007; Scriven, 2007). This means that if a given checklist is consistently used, other evaluators could replicate a given evaluation based on the same framework.

In addition to the features specifically highlighted in the literature, evaluation checklists also support metaevaluation, during which the quality of the evaluation work and products can be assessed. Five types of checklists have been put forth by Scriven (2007), including laundry lists, strongly and weakly sequential checklists, iterative lists, diagnostic checklists, and criteria of merit (COM) checklists. Characteristics and examples of each type of checklist are displayed in Table 7. In addition to the examples provided in the table, others are available in varying professional environments; including medicine, psychology, aeronautics, education, academe, and evaluation, among many others (Hales et al., 2007).

Evaluation checklists have been around almost as long as the evaluation profession has been growing in the United States. Stufflebeam (2001a), one of the leading figures in evaluation checklist development, reflects on his first experiences with evaluation checklists during the late 1960s. In this reflection, he notes that his students at Ohio State University repeatedly requested a protocol for planning evaluation that, at that time, was not readily available. In response, Stufflebeam developed his first checklist comprised of a list of questions that were categorized in
six sections (focusing the evaluation, collecting the needed information, organizing the information, analyzing the information, reporting the findings, and administering the evaluation). This checklist is now available on The Evaluation Center’s (EC) Checklist Project Web site as a revised and improved version titled “Evaluation Plans and Operations Checklist.” Initially developed on his experience interacting with evaluation stakeholders, Stufflebeam (2001a) states:

I doubt that I have ever tried to pose these questions to clients or other stakeholders in any particular sequence. Moreover, I didn’t ask clients to answer particular questions if I had gotten the information simply by listening and reading materials . . . . I also observed that the checklist probably wouldn’t work if they [Stufflebeam’s students] treated it as a rigid, linear protocol for interviewing clients (p. 73).

Table 7

A Taxonomy of Checklist Types (based on Scriven, 2007)

<table>
<thead>
<tr>
<th>Type</th>
<th>Characteristics</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laundry list</td>
<td>Validity is not impacted by the ordering, but could be affected if items are incorrectly categorized or grouped.</td>
<td>Travel lists</td>
</tr>
<tr>
<td>Strongly/weakly sequential</td>
<td>Validity is impacted by the ordering, categorization, and grouping of items.</td>
<td>Flight crew checklist⁴⁷</td>
</tr>
<tr>
<td>Iterative</td>
<td>Validity is enhancing via repeated review of checkpoints. Consideration of latter checkpoints may require a review and adjustments of earlier ones.</td>
<td>Medical procedures lists (e.g., continuous and repeated measures of vital information of a patient during surgery)</td>
</tr>
</tbody>
</table>

⁴⁵ Cf., http://www.wmich.edu/evalctr/checklists/checklistmenu.htm
⁴⁶ For examples specific to medical checklists, see Hales et al. (2007).
Table 7—continued

<table>
<thead>
<tr>
<th>Type</th>
<th>Characteristics</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic</td>
<td>These are based on flowcharts to yield broad conclusions.</td>
<td>Some lists in psychological assessment (e.g., in the Diagnostic and Statistical Manual [DSM-IV])</td>
</tr>
<tr>
<td>COMlist</td>
<td>Order, completeness, categorization, and grouping affect validity of evaluative conclusions.</td>
<td>Rating sheets in contests or for job applicants</td>
</tr>
</tbody>
</table>

Under the auspices of the EC Checklist Project, many other evaluation checklists have been developed, including several others by Stufflebeam, as well as key evaluation figures such as Michael Scriven, Michael Quinn Patton, Robert Stake, Egon Guba and Yvonna Lincoln, and Ernest House and Kenneth Howe. At this time, the EC Checklists Web site includes thirty-one lists that are intended to support evaluation students, practitioners, funders, and other stakeholder groups in planning, implementing, designing, managing, creating, and evaluating evaluations.

Why Develop Evaluation Checklists?

While the Checklists Web site is an abundant source for evaluation checklists, there is little published research on how these were validated and used. The only account of research or evaluation on the checklists is Bichelmeyer’s (2002) study of usability attributes related to the Web site’s checklists content, format, titles, and accessibility. However, research on checklist content and use is beneficial and available in other disciplines (Manley & Cuddeford, 1996; Wolff et al., 2004; Hales et al., 2007).
For example, Hales and colleagues (2007) discuss evidence from checklist use that confirms performance improvement, minimization and prevention of errors, and improved task management, especially in the aeronautical and medical sectors (see also Wolff et al., 2004). In essence, research has shown that by condensing large amounts of information, checklists increase reliability and validity in performing a given task and facilitate understanding, thus enable use and task performance (Hales et al., 2007), while at the same time reducing the influence of *halo*\(^{48}\) and *Rorschach*\(^{49}\) effects (Scriven, 2007). Other strengths of checklists that have been pinpointed by Persaud (2007) include their potential for minimizing the amount of time involved in assuring that the task is done well, thus reducing monetary cost associated with carrying out a given task.

Hales and colleagues (2007) also emphasize that negative impacts from checklist use have not yet been illuminated in the published literature and that research has not found negative effects of checklist use (Manley & Cuddeford, 1996). However, potential negative impacts may include overuse, related fatigue, and unnecessary barriers (e.g., such as adding unnecessary details) in performing given tasks resulting from needless complexities within given checklists. Persaud (2007) lists additional shortcomings and potential pitfalls of evaluation checklists, including:

\(^{48}\) The *halo* effect, similar to the *devil* effect, involves highly positive judgment about an evaluation object or a person based on a general overall impression or a specific impression on one particularly positive trait. The *devil* effect yields negative judgments based on one negative judgment based on one critical dimension (cf., Thorndike, 1920).

\(^{49}\) The *Rorschach* effect occurs when the evaluator sees what he wants to see, whereas checklists facilitate taking all important aspects into account.
(1) Creation of tunnel vision and reduced user creativity: That is, users may follow the checklist too exactly without thinking creatively about its shortcomings and alternative uses.

(2) Naive, nonreflective use: Checklists are not perfect and may include outdated information as well as mistakes (even if the tool has been reviewed and revised). If flaws cannot be identified by users, results of checklist use could be inaccurate, thus negatively impacting decision making.

(3) While checklists incorporate large amounts of knowledge about the area under investigation, they are no substitute for formal training. For most effective checklist use, training is required about the checklist’s content. However, checklists provide viable strategies (see also Sanders, 2001) for users to enhance task performance and minimize errors of omission and commission.

What Other Sustainability Checklists are in Existence?

Several sustainability and sustainability evaluation checklists exist. Because many are components within larger organizational frameworks (e.g., ILO’s PARDEV, as shown in the table below), their dissemination and availability are limited. However, a few examples of sustainability checklists can be found via the Web, in the gray literature50, and in traditionally published texts. As indicated in Table 8, these checklists usually focus on a very specific issue of sustainability evaluation, specify key criteria for sustainability, or emphasize a certain aspect of sustainability.

---

50 Gray literature is text that is not easily to be found. It may include unpublished reports, research papers, and other documentation.
evaluation, but none of these checklists considers both evaluation of sustainability and evaluation for sustainability.

Table 8

**Sustainability Checklists: A Few Examples**

<table>
<thead>
<tr>
<th>Checklist title, source, and accessibility</th>
<th>Brief description</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sustainability Checklist: Part of the Sustainable Development &amp; Construction SPD</em> (Guildford Borough, 2005), www</td>
<td>This checklist is part of an application for approval of development or construction of a residential unit or commercial floor space, including three sections (one for both types of applicants and one for each type of applicant) with a total of thirty-four questions to be answered as yes/no. Point of specific consideration include the local context, the built form, energy, water, movement, waste and recycling, landscape and wildlife, heritage, landscape, and sense of place.</td>
<td>Environmental sustainability</td>
</tr>
<tr>
<td><em>Sustainability Checklist</em> (Lawrenz &amp; Keiser, n.d.), www</td>
<td>Created for projects funded by NSF’s Advanced Technological Education program, this evaluation sustainability/checklist emphasizes sustainability via external support mechanisms and institutionalization.</td>
<td>Project/program sustainability/ continuation</td>
</tr>
<tr>
<td><em>Sustainable Checklist for LEED</em> Projects Sustainable Sites* (Los Angeles Community College District Proposition A/AA Bond Program, 2004), www</td>
<td>This is an accountability tool intended to assist architects and planners to comply with minimum standards in renovation projects. Users must indicate whether items on the list were incorporated in the design, the cost of impact, and effects on the schedule.</td>
<td>Environmental sustainability</td>
</tr>
<tr>
<td><em>Sustainability Checklist</em> (International Labor Organisation’s [ILO] Department of Partnerships and Development Cooperation [PARDEV], n.d.), Personal communication/ Gray literature</td>
<td>This checklist specifically considers sustainability at the outset of a project by asking nineteen questions related to political support, institutional capacity, sociocultural issues, gender issues, appropriate technology, protection of the environment, and economic and financial issues.</td>
<td>Sustainable development checklist</td>
</tr>
</tbody>
</table>

51 Accessibility in the table header refers to where the example could be found.
52 Leadership in Energy and Environmental Design
Table 8—continued

<table>
<thead>
<tr>
<th>Checklist title, source, and accessibility</th>
<th>Brief description</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sustainability Evaluation</strong> (Stufflebeam, 2007b), www</td>
<td>Sustainability evaluation comprises the seventh component in the CIPP model and considers institutionalization of a program’s contribution over time. The checklist emphasizes activities to be implemented by the evaluator and clients/stakeholders.</td>
<td>Project/program sustainability/ continuation</td>
</tr>
<tr>
<td><strong>Checklist for Assessment of Projects</strong> (Thierstein &amp; Walser, 2007), book</td>
<td>This quantitative weight-and-sum checklist enumerates ten criteria and forty-eight subcriteria. The main criteria include (1) economy, (2) ecology, (3) society and societal matters, (4) forms of participation, (5) form of networking, (6) subsidiary in interaction between different institutional levels, (7) diversity in interaction between different approaches, (8) equity between groups, (9) equity between regions, and (10) equity between generations. It is argued that the five overarching dimensions—(i) economy, (ii) ecology, (iii) society and societal matters, (iv) forms of participation and networking, subsidiary in interaction between different institutional levels and diversity in interaction between different approaches; and (v) equity between groups, regions, and generations—are equally weighted. Yet, dimensions vary in their number of indicators (between eight and twelve), all of which are to be summed at the end of each dimension.</td>
<td>Sustainable development evaluation</td>
</tr>
<tr>
<td><strong>An Index for Assessing Project Sustainability</strong> (Valadez &amp; Bamberger, 1994), Personal communication/ Gray literature</td>
<td>This is a four-dimensional rating sheet with specific consideration of continued delivery of services and benefits, maintenance of physical infrastructure, long-term institutional capacity, and support from key stakeholders</td>
<td>Project/program sustainability/ continuation</td>
</tr>
</tbody>
</table>
| **Sustainability Checklist** (Woking Borough Council, 2006/07), www | This sustainable development checklist has eighteen dimensions that were developed based on Agenda 21 and include the following considerations:  
  • Minimization of resource use and pollution  
  • Accentuation of biodiversity, access to the natural environment, and features that contribute to Woking’s pride of place | Sustainable development checklist |
Table 8—continued

<table>
<thead>
<tr>
<th>Checklist title, source, and accessibility</th>
<th>Brief description</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Encouragement of a strong and diverse local economy, employment opportunities with good working conditions, opportunities for education and information, personal safety and property security, opportunities to be part of the local community, and opportunities for decision making</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Meeting local needs locally</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Promotion of good physical and mental health and treat poor health, equality in health, and in equality in general</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Accessibility to affordable, balanced diet; transport that does not rely on the car; and informal leisure opportunities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provision of appropriate housing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Sustainability Tool-Kit* (Ridde, Pluye, & Johnson-Lafleur, 2007), www

This toolkit consists of two tools that differentiate between sustainability processes and sustainability level of projects. Each tool provides means for collecting and synthesizing data. In terms of the sustainability process specifically, different types of events are taken into account. To determine the sustainability level, a list of indicators categorized within five criteria (memory, adaptation, values, rules, and institutional standards) is taken into account.

**Checklist Development Methodology**

In this section, procedures used for developing checklists are described. Different models for checklist development are discussed comparatively, considerations for formatting checklists are highlighted, and steps involved in developing the sustainability evaluation checklist are presented.
Models for Checklist Development

Two models for checklist development were considered for the construction of the SEC: (1) Stufflebeam’s (2000a) Checklist Development Checklist and (2) Hales and colleagues’ (2007) Key lessons learned for medical checklist development. As the title of Stufflebeam’s (2000a) work suggests, his Checklist Development Checklist is relatively generic and could be applied to various contexts in which checklists are used. In contrast, the model presented by Hales et al. (2007) is specifically keyed to medical checklists. Figure 8 contrasts both models and relates them back to this dissertation’s content.

Comparing both lists suggest some overlap as well as possible discrepancies. Convergences are illustrated in Figure 8 via two-sided arrows and correspond to the overall flow of the dissertation. That is, Chapter 1 focused the checklist task and described the need for the SEC, as well as its potential users and uses. Chapter 2 provided the frame for developing the checklist content, which will be elaborated on later in this chapter. The current chapter will also discuss initial review strategies. Chapters 4 and 5 detail the evaluation methodology and results from the checklist review, and Chapter 6 finalizes the checklist for the purposes of this dissertation and provides future direction on continuously improving the checklist.
**Figure 8. Models for Checklist Development**
For now, the following discrepancies are of interest. First, and because the Hales model was created specifically for medical checklist development, it includes important details that may not necessarily apply to the international development evaluation context. These details are inherent in points 4 and 7 in Hales and colleagues' model. In essence, under point 4, Hales and colleagues stress the importance of the "flow of real-time user activities" and the "clinician state of mind." While the authors fall short in elaborating on the distinct meaning of the items, it appears commonsense that the "flow of real-time user activities" is not critical in most evaluation contexts. An exception, however, would be a COMlist used for rating real-time performances.

The second discrepancy of importance here pertains to point 7 in the Hales model, specifically referring to obtaining authorization from regulatory powers. The evaluation checklist discussed herein, as well as the checklists available on the EC Web site, assume adoption at one's own risk. That is, obtaining approval for using a checklist in a given context is not the responsibility of the checklist author. Instead the user is responsible for acquiring approval of the methodologies used by the client.

One element receiving distinct attention by both authors and others is the format or design of a checklist. The next subsection, therefore, emphasizes considerations for formatting checklists.
Considerations for Formatting Checklists

Formatting in evaluation plays an important role for enhancing understanding of findings from evaluations, increasing evaluation use, and structuring mnemonic devices such as evaluation checklists (Bichelmeyer, 2002; Hales et al., 2007; Lawrenz, Gullickson, & Toal, 2007; Stufflebeam, 2000a; Torres, Preskill, & Pionek, 1997) and usually relates to the way in which a document or presentation is organized and presented.

Stufflebeam (2000a), Bichelmeyer (2002, 2003), and Hales and colleagues (2007) provide specific sets of considerations for formatting checklists. Stufflebeam (2000a) emphasizes formatting considerations (see point 8 in Figure 8) in relation to evaluative operations that were discussed in Chapter 2 of this dissertation. That is, formatting decisions are based on the need for weighting, scoring, or profiling requirements. Examples of checklists formats based on evaluative operations can be found on the EC’s Checklists Web site, specifically within the section on metaevaluation checklists (see, for example, Stufflebeam 1999c, 2000b, 2000c).

Bichelmeyer (2003) developed a Checklist for Formatting Checklists (CFC) which is divided into five primary categories. Instead of only considering evaluative operations in making decisions about the checklist format, Bichelmeyer promotes consideration of context, content, structure, inclusion of images, and usability. Both Stufflebeam’s and Bichelmeyer’s considerations were taken into account in Hales and colleagues’ (2007, p. 25) formatting guidelines for medical checklists.
Procedures Used for Developing the SEC

How was the SEC developed? First, a comprehensive review of the literature helped to focus the checklist's content, provided references for intended uses and users, and revealed a variety of potential checkpoints related to evaluation in general and sustainability evaluation specifically (see Chapter 2). Initial ideas pertaining to the checklist were presented during the “Evaluation of Sustainability: European Conferences and Training Courses” (EASY-ECO) conference in Saarbrücken, Germany, and the biennial conference of the European Evaluation Society in London, United Kingdom, in 2006 (see Schröter, 2006a, b). With preliminary feedback from the conferences in mind, drafts of the checklist (2006, 2007) were created and refined, and then revised after further feedback was obtained during Evaluation Café discussion forums at Western Michigan University’s Evaluation Center (Schröter, 2007a, 2008a), and presentations at the annual conference of the American Evaluation Association in Baltimore, MD (Schröter, 2007b, 2007c). Each presentation focused on different elements or stages of checklist development and yielded critical feedback that was incorporated into the test version of the checklist, which is discussed at the end of this chapter and included in Appendix B.

The feedback received was limited in terms of the checklist’s content, and much advice was targeted toward the checklist’s format. In essence, most reviewers criticized the length of the original checklist, which included multiple figures and tables that—in the author’s opinion—would facilitate the evaluation task. In response and to warrant an adequate sample size for the study presented in Chapters 4 and 5 of
this dissertation, the document was shortened significantly, redundant information was removed, and the language was refined and simplified for the intended audiences; a glossary of terms was added in an appendix to facilitate understanding. The content of the test version of the checklist is further elaborated in the subsequent section of this chapter.

The Sustainability Evaluation Checklist

The checklist described has characteristics of various checklist types that were introduced by Scriven (2007). It is an iterative, yet weakly sequential list that incorporates features of a laundry list (i.e., core tasks to be considered when conducted evaluation) and criteria of merit, worth, and significance (i.e., criteria of specific relevance to sustainability). As an iterative list, the SEC requires the user to consider each checkpoint more than once to assure that everything is adequately addressed. Yet, the SEC has a weakly sequential structure, beginning with general considerations about the evaluation and the evaluand and later emphasizing specific criteria of merit, worth, and significance that are essential to the evaluation of/for sustainability. To some degree the SEC is a laundry list, because it attempts to capture everything needed to conduct sound sustainability evaluations by reducing errors of omission and commission, increasing the usefulness of the evaluation, and reducing the influence of memorable observations of one dimension over others. The core of the checklist comprises the criteria checkpoints, presented in three separate parts that
explicitly differentiate criteria of merit, criteria of worth, and criteria of significance. The following subsection summarizes key concepts derived from the literature review as well as the core content of the checklist.

**Key Concepts**

The SEC is intended for use in evaluating the sustainability of social and international development programs and projects. Within the checklist, sustainability is broadly defined as the *continuation or maintenance of human, social, economic, and/or environmental benefits (processes or outcomes) from projects or programs.* Evaluation is defined as the determination of merit (quality), worth (value), and significance (importance).

Two approaches to evaluating sustainability are differentiated within the SEC. First, as a moral principle or global goal requiring a project or program to continually improve its processes within the constraints of the ecological environment and without producing harm to human, social, economic, or environmental dimensions, sustainability is a process-oriented, prospective concept that must be considered throughout a program’s life cycle. Second, as an outcome or impact from programmatic efforts, sustainability is a retrospective concept in which positive intended and unintended outcomes and impacts are maintained beyond the removal of initial program funding. Such outcomes include continuation, replication, and exportability (i.e., generalizations) of program activities and impacts.
The SEC is intended as a guide for both external and internal evaluation practitioners. However, evaluation funders and program managers may also benefit from the extensive nature of the checklist, because it includes many aspects of relevance in program planning and evaluation. Experienced evaluators will benefit from the comprehensive nature of the checklist, while new evaluators or those with limited exposure to sustainability evaluation will find the checklist instructional. As a framework for planning and designing evaluation in various contexts, the checklist does not dictate exactly how each application of each checkpoint should be executed. However, it encourages the user to consider the relevant aspects in sustainability evaluation. The checklist does not promote any specific research methodology; it only alludes to key approaches. It is the responsibility of the user to determine what methods and approaches are most appropriate in collecting and analyzing data about evaluands' sustainability. Data collection and analysis plans and instruments have to be developed in relation to the specific task at hand and for the individual evaluand.

Checklist Content

Figure 9 illustrates the flow of the checklist. As shown, the checklist begins with an introduction that clarifies its purpose, intended users, key characteristics, and definitions and descriptions of key concepts (i.e., sustainability, evaluation, evaluation of sustainability, evaluation for sustainability, and checklist). Throughout the checklist, asterisks (*) are used to indicate terms that are explained in a glossary that is appended to the checklist.
The SEC's overarching purpose is to guide practitioners in planning and designing evaluations of sustainability for sustainability. Evaluation of sustainability takes a retrospective look at sustainability, thus it specifically considers the extent to which the evaluand and its outcomes have been maintained or continued to date. Evaluation for sustainability embraces the forward-looking thinking of the sustainable development community and is specifically concerned with the extent to which the evaluand prospectively contributes to improving human, social, and economic conditions under consideration of ecological confines.

As shown in Figure 9, the checklist is divided into two parts and six sections, which are built on Scriven's (1980, 1995) and Fournier's (1995) logic of evaluation. In the first part (sections 1-3), general considerations of conducting evaluation are discussed. The second part (sections 4-6) emphasizes criteria of merit, worth, and significance.

The rationale for the inclusion of each checkpoint is discussed below.

**Part A: General considerations in evaluation.** This first part includes three subsections which provide the overarching frame for an evaluation and should be discussed among evaluation team members, the evaluation client, and key evaluation stakeholders to clarify information needs, resources, methodological decisions, and required levels of detail—as well as to inform evaluation management. Although the components discussed in this section comprise considerations that are of general concern in evaluative endeavors, the SEC indicates special concerns for sustainability evaluation. The rationale for including the general considerations is based in the
following assumptions about checklist users: (i) they may not necessarily be aware of the pertinent evaluation literature; (ii) they may not have (easy) access to readily available checklists or similar tools (i.e., poor or no Internet connection); (iii) they may not be aware of the existence of evaluation checklists.

**Figure 9. Checklist Flow**

Because the WMU EC checklists are freely available and accessible via the Web, the SEC recommends accessing these checklists as supplementary resources, but attempts to capture their key tasks to minimize the need for users to search out
additional resources for planning and implementing evaluations. Details on choice of methodology or evaluation approaches are not included in the checklist to maintain focus on the core tasks and to avoid methodological debates.

The first section under the general considerations includes eight checkpoints, each of which is comprised of questions or comments intended to ignite further thought about these issues (c.f., Beck, 2006; Gibson et al., 2005; Greene, 2005; Hardi, 2007; Joint Committee on Standards for Educational Evaluation, 1994; ODI, 2006; Patton, 2008; Scriven, 1991):

1. Evaluation purposes
2. Differing directions for initiating evaluations
3. Evaluator roles
4. Potential evaluation audiences
5. Timeframes under consideration of life cycle conceptions
6. Types of evaluation (without skewing the user toward specific evaluation models or approaches)
7. Key evaluation questions
8. Metaevaluation

Issues of specific relevance to sustainability evaluation are highlighted in places where it was perceived as important. For example, in determining the direction of the evaluation (point 2 above), it appears common that top-down evaluation work is usually concerned with sustainability as one dimension among many others (e.g., OECD, 1992; ODI, 2006).
The timeframe under evaluation (point 5) is illuminated in terms of prospective and retrospective considerations for the evaluation of sustainability for sustainability. In essence, at the program development stage, sustainability is considered prospectively (European Commission Budget, 2001; GAO, 1990; World Bank IEG, 2007). During program implementation and maturation, levels and breadth of sustainability are considered. Once initial resources are reduced or terminated, the stability of the evaluand—as well as the quality of outcomes and impacts—are considered in terms of reduced resources or altered funding streams (e.g., Welch & Gullickson, 2006).

When considering key questions to be answered by the evaluation (point 7), distinctions between evaluation of and for sustainability must be clear, as well as whether the evaluation is intended to make conclusions about relative or absolute merit of sustainability or for sustainability (Davidson, 2005a).

Once the user clarifies all general issues of relevance to the evaluation, the emphasis shifts to the evaluand, section 2. This section consists of four checkpoints intended to increase understanding and awareness about the evaluand, including consideration of these four questions:

1. What exactly is evaluated?
2. Who is involved and impacted?
3. What is the potential reach of the evaluand?
4. What are the components, dimensions, and potentially theoretical assumptions of the evaluand?
This section is not only intended to increase an understanding of the evaluand, but also the systemic nature of policies and programmatic intervention. Moreover, the phenomenon of sustainability is clarified within the specific context of the evaluand under investigation.

Section 3 adumbrates the four-step general logic of evaluation and provides guidance in developing the working logic for sustainability evaluation (Fournier, 1995; House, 1995; Rog, 1995; Scriven 1980, 1991, 1995, 2005):

1. Identification of values/criteria
2. Standard setting
3. Data collection
4. Synthesis

To support identification of criteria, reference is made to various sources for criteria in general and to pertinent sustainable development and international development policies specifically (NSDS, PRSPs, MDGs, UNDAF, Agenda 21). The standard-setting procedures (Coryn, 2007; Scriven, 2005) are kept general to allow the user to establish standards appropriate to the specific context. For data collection purposes, special reference is made to available indicator sets (e.g., Bossel, 1999; Meadows, 1998; Pinter, Hardi, & Bartelmus, 2005; United Nations Commission, 2001), time and space considerations and broad participation (e.g., Hardi & Zdan, 1997). Data collection methods and data-analytic techniques are not explicated, assuming that users have these research skills, but lack specific proficiency and education in evaluation-specific logic and methodology. Similar to the data collection
component, the synthesis checkpoint is treated in general (Fournier, 1995; Scriven, 1995, 2005).

**Part B: Criteria for evaluating sustainability.** The second part of the checklist focuses on criteria of merit, worth, and significance for sustainability evaluations. The distinction between criteria that address questions of merit, worth, and significance separately is new and based in the analogous reasoning with the definition of evaluation ala the determination of merit, worth, and significance. Separate contemplation is intended to enable checklist users to consider all three types of evaluation independently. Usually, criteria of merit lists (cf., Scriven 2006, 2007) are put forth, while considerations of significance are either embedded within the criteria of merit or placed within comparison and generalizability consideration (specifically see Scriven, 2006). Criteria of worth are usually considered in specific cost checkpoints (e.g., Scriven, 2006) or whole cost analysis checklists (e.g., Persaud, 2007; Smith, 1983). Criteria of significance are further defined as those properties of an evaluation object that are part of good sustainability in a given context.

Within section 4, which focuses on criteria of merit, a distinction is made between process-orientated and outcome-oriented criteria. Process-oriented criteria include those elements that were identified as especially useful in determining if the evaluand has the capacity for addressing sustainability needs (i.e., prospective considerations) and which of the evaluand’s elements (e.g., activities) persist after initial external support for the evaluand have been removed (i.e., retrospective
considerations) or complemented with other resources. There are seven process-oriented checkpoints:

1. Leadership (e.g., Trzyna, 1995; Fullan, 2005)
2. Organizational characteristics (e.g., Howard & Howard, 2000; Pezzey, 1992)
3. Infrastructure (e.g., Goodland, 2002; Hardi, 2007; Johnson et al., 2004)
4. Collaboration (e.g., Aaker & Shumaker, 1996; Hardi & Zdan, 1997; Mancini & Marek, 2004)
5. Understanding the community (e.g., Larsen, 1999)
6. Responsiveness (e.g., Miller, Kobayashi, & Noble, 2006)
7. Use of evidence from monitoring and evaluation (Gibson et al., 2005).

Outcome-oriented criteria are those relevant to determining if the evaluand has the capacity for sustainability (i.e., prospective considerations) and/or which outcomes have been sustained to date (i.e., retrospective considerations). Two checkpoints are concerned with outcome criteria:

1. Goal orientation in terms of global issues of importance in sustainable development (e.g., the systemic nature of human, social, economic, and environmental impacts), (Goodland, 2002)
2. The evaluand’s positive and negative impacts regarding the four pillars of sustainable development, as well as cross-dimensional impacts
The section on criteria of merit is followed by a section on criteria of worth. While available cost evaluation checklists provide guidelines for identifying and analyzing costs and benefits that can be useful in many evaluations (e.g., Smith, 1983; Persaud, 2007), the SEC emphasizes the need to consider sustainability and inherent cost assumptions in terms of current and future generations. Five checkpoints are put forth:

1. **Time at which costs/resources are accrued**
2. **Stakeholders and impactees to whom costs accrue**
3. **Types of costs, including actual costs to humans, society, global costs, and opportunity costs**
4. **Specific costs and resources use to consider**
5. **Risks as a special type of costs**

Finally, section 6 discusses criteria of significance, those that provide information about the importance of sustainability in a given context. Two overarching categories are included:

1. **Needs of impactees**
2. **Spectrum**

Within the needs cluster, distinctions are made between human, social, economic, and environmental considerations. Under spectrum, the scope and the scale of the intervention are emphasized.

The checklist concludes with a glossary of terms and a list of references used for developing the checklist to promote further reading on the subject.
Summary

This chapter provided background information about checklists in general and checklist development specifically. It also discussed procedures involved in developing the SEC, its key concepts, and a summary of the SEC’s components. At this point, the SEC is available only in static formats (MS Word, Rich Text, and PDF). However, to achieve full functionality and utility, the information in the SEC are to be interlinked in the future for most efficient use. Yet, prior to disseminating the tool widely, extensive reviews have to take place. In essence, the checklist and specifically the criteria checkpoints raise some uncertainties, although the list appears comprehensive at first sight. Thus, the subsequent chapters emphasize means for checklist evaluation, a discussion of the study’s results, and future implications for checklist development in general and the SEC specifically.
IV. METHODOLOGY

The previous chapter discussed procedures involved in developing the SEC and elaborated on its contents and rationale. This chapter describes the methodology for the evaluation of the checklist by experts and potential users. Specifically, the purpose, participants, methods and materials, and research procedures are discussed.

An overview is provided in Figure 10.

**Figure 10. Chapter IV Overview**
Purpose

The purpose of this nonexperimental, exploratory, primarily qualitative study was to garner critical, detailed feedback on the SEC in varying development contexts via interviews and questionnaires. A qualitative, descriptive design was chosen to validate the concepts within the checklist, because the SEC is a new evaluation tool that incorporates concepts that are widely discussed among evaluators and practitioners in the sustainable development arena (Hoepfl, 1997). The study used Stufflebeam’s (2000a) and Hales and colleagues’ (2007) suggestions for checklist evaluation and incorporated (1) an expert and practitioner review, (2) application of the checklist by intended users, and (3) a synthesis of the information from both the initial appraisals and experiences from field tests. The overarching questions, listed below, were intended to yield in-depth information about the tool and its potential for improvement:

1. What is missing from the checklist?
2. What components or checkpoints are not necessary?
3. Are there any other errors or problems that need to be addressed?
4. What, if anything, did you like about the checklist?
5. Do you have any suggestions for how to improve the checklist?

In essence, the study was designed to explore the extent to which the SEC had content validity and was perceived as useful in its draft stage. In this context, validity was conceptualized broadly as the checklist’s potential for guiding objective and
comprehensive consideration of key aspects of sustainability within evaluations of national and international development interventions (see Trochim, 2006). While conceptualizations of validity usually refer to empirical measures (Neuendorf, 2002), evaluation checklists are unique in that they are not intended for direct assessment purposes, but rather as tools that guide evaluators through a set of tasks and considerations when planning an evaluation and developing related assessment tools.

Messick's (1989) validity framework was used as a basis for conceptualizing the study, where validity is defined as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on" (p. 13) a given device or tool. More specifically, Messick argues that validity is "an inductive summary of both the existing evidence for and the potential consequences of score interpretation and use" (p. 13), where scores to refer broadly to quantitative or qualitative summaries of conclusions based on the use of a given instrument.

In terms of the SEC, the primary question was: What empirical evidence is useful for validating a framework which by itself is not a traditional instrument and neither has been implemented to its full extent, but rather comprises a proposition for conducting evaluations of a widely discussed phenomenon that comprises a criterion within some evaluation, yet a concept that is also being evaluated on its own right? Usually, evaluation checklists are validated and refined by their usage over long periods of time and have credibility because of the extensive experience, knowledge, and professional status of their authors. However, the SEC is a new tool, and the
study described here comprises early efforts to evaluate the checklist prior to encouraging its use.

Scriven (2007) and Stufflebeam (2000) provide some sets of criteria for exploring the goodness and usefulness of evaluation checklists. Scriven (2007) emphasizes the following characteristics:

a) Checkpoints should have criterial status, that is, be comprised of dimensions of goodness rather than indicators.

b) Lists should be complete and concise.

c) In cases where the checklists are intended for scoring purposes, items should not be overlapping.

d) Criteria within checklists should be commensurable, clear, and confirmable.

Stufflebeam (2000) suggests that checklists should be investigated in terms of their "pertinence, comprehensiveness, clarity, applicability to the full range of intended uses, concreteness, parsimony, ease of use, and fairness" (p.3).

Comparing the criteria set forth by both experts suggests that good evaluation checklists are characterized by reducing omissions, commissions, and other errors in doing evaluation. Additionally, checklists should be complete (Scriven, 2007) or, in Stufflebeam's (2000) terms, comprehensive. Scriven suggests that criteria within checklists must have criterial status, while Stufflebeam emphasizes pertinence.

---

53 Criteria provide the dimensions on which the goodness of the checklist can be judged.
Scriven recommends conciseness and Stufflebeam calls for parsimony, both concepts relating to prudence and brevity. Both authors agree that checkpoints must be clear.

The remaining criteria suggested by the two experts are more distinct. Scriven discusses the commensurability of criteria, suggesting equal levels of generality. He also calls for criteria to be observable (i.e., confirmable). Stufflebeam, on the other hand, emphasizes ease of use and applicability to the full range of intended uses, specifically emphasizing utility, but also fairness.

In sum, to have validity, the checklist must at least fulfill the following requirements:

a) It must be complete; meaning no essential elements for the evaluation of sustainability for sustainability should be missing.

b) It should not include redundant or unnecessary information (i.e., commissions).

c) The checkpoints should be independent from one another.

d) The criteria within the checklist must exemplify aspects of good, valuable, and important sustainability without tapping into indicators.

e) The checkpoints should be clear and concise.

f) Overall, the checklist should be useful for intended uses by intended users.

To explore alternatives for assessing the validity of the SEC, Messick’s (1995) unified validity framework has been considered (see Figure 11).
Figure 11. Messick’s (1995) Unified Validity Framework

As shown, the framework represents two interconnected aspects of validity. The first aspect is the source for justifying sustainability evaluation with the help of the checklist based on evidence or consequences. The second aspect comprises checklist interpretation and use. As illustrated, construct validity permeates each cell. As interpretive meaningfulness, construct validity subsumes varying types of validity, including content validity (Messick, 1980, 1981, 1989), which is the focus of the study described here. The content component of the larger concept of construct validity comprises the degree to which the SEC is relevant to and representative of the topic under investigation (Colton & Covert, 2007; Messick, 1980, 1989, 1995). It is established by showing that the checkpoints correspond to the “universe of
sustainability evaluation" (cf., Cronbach & Meehl, 1955). Content validation also provides evidence about "usefulness, domain, facets, boundaries, and predictive efficacy of the construct" (Haynes et al., 1995, p. 239). Because the validation process is described as especially difficult for ambiguous constructs (like sustainability), alterations, refinements, and reexaminations of the checklist were anticipated and planned.

Messick's unified framework suggests three additional elements for consideration in validity studies. As shown in cell 2, Messick (1995) adds relevance and utility to the evidential basis for checklist use, meaning that in addition to general validity considerations, the benefits of using the SEC should be considered. Value implications for consideration within the consequential basis of interpretation of the SEC are added in cell 3, and suggest linkages between checklist interpretations and applied practice and policy. This is especially relevant when considering multicultural validity of the checklist, which is defined as the extent to which the checklist "measures what it purports to measure as understood by an audience of a particular culture" (Colton & Covert, 2007, p. 69). Thus, the study targeted multicultural, international respondents. Finally, cell 4 subsumes construct validity, relevance and utility, and value implications—plus social consequences, which refer to "potential and actual" impacts and side effects that result from using the checklist (Messick, 1995).

\[54\] Usually in relation to costs
In sum, this study was intended to explore the validity of the checklist by considering criteria put forth by Stufflebeam (2000) and Scriven (2007), as well as elements from the pertinent validity literature. As a result, the checklist’s content, utility, and potential consequences were explored in the study.

Participants

Participants included evaluation, international development, and sustainable development experts and practitioners. The following section describes the procedures used for sampling, recruitment, and protection of human subjects’ rights and welfare.

Sampling Procedures

Purposeful sampling, specifically maximum variation (heterogeneity) sampling, was used to select potential interviewees with significant knowledge and experience in evaluation, development, and/or sustainability theory. Evaluation and sustainable development experts were deemed appropriate for answering the questions posed in this study because of their extensive knowledge in the domains under consideration and the strong focus of the interviews on the checklist’s content (Flick, 2006a, 2006b). Selection criteria included (i) contributions to the literature,

55 A strategy intended to capitalize on the information gained by considering varying perspectives (see Patton, 2002, pp. 234-235)
discussions, and practice on sustainability; and/or (ii) expertise in evaluation or
development; and (iii) availability of contact information (i.e., e-mail addresses).

Purposeful, convenience sampling procedures were employed to reach
potential respondents. The sampling was purposeful in that potential SEC users were
targeted by inviting members of relevant professional listservs. The sample was also
convenient, because the level of experience with evaluation and specifically with
evaluation of sustainability could not be established a priori. In essence, any member
of the listserv with an interest in the study could self-select to participate. However,
the sampling frame was purposefully chosen to gain feedback from individuals
concerned with evaluating sustainability in different development contexts in various
countries and regions around the globe. Practitioners were expected to have insights
about the SEC’s relevance, usefulness, and validity in diverse contexts and cultures.
Additional individuals were added to the original list based on references from
respondents, adding a snowball sample to the design.

Recruitment

Using the mail merge function in Microsoft Word, a personalized e-mail was
sent to the pre-identified expert interviewees on March 30, 2008. Within the e-mail
(see Appendix C), the nature and purpose of the project were explained and experts
were asked if they had the time and interest to contribute to the study.

Because it is well known that experts have extreme time constraints, they
were provided with the options to respond to the core questions via e-mail, participate
in the online survey, or schedule an interview. After a positive response was returned, each expert was contacted individually with the draft SEC and detailed information about the task (see Appendix D).

*Questionnaire participants* were recruited via messages (see Appendix E) posted to professional listservs. Specifically, invitations were submitted to members of MandENEWS,56 XCeval,57 and the Pelican Initiative.58 MandENEWS is a listserv with more than 1,700 members who have an interest in monitoring and evaluation of international development aid. XCeval, originated by the American Evaluation Association’s (AEA) International and Cross-Cultural topical interest group (TIG), is a listserv with more than 600 members that encourages exchange of ideas related to international development evaluation. The Pelican Initiative is a medium to exchange ideas about evidence-based learning and communications to contribute to social change. The platform has more than 400 members including international development practitioners and policymakers with various disciplinary backgrounds and specialties. The invitation was also extended to several other development group (Dgroup59) administrators and individuals who gave presentations in International and Cross-Cultural TIG-sponsored sessions at the 2007 AEA conference. Furthermore, interested parties were encouraged to distribute the invitation to others who may be interested in the study. As a result, the invitations were posted to the Washington

---

56 http://tech.groups.yahoo.com/group/MandENEWS/
57 http://groups.yahoo.com/group/XCeval/
58 http://www.dgroups.org/groups/pelican/index.cfm
59 http://www.dgroups.org/about/
Protection of Human Subjects

Prior to initiating the study, Western Michigan University (WMU) Human Subject Institutional Review Board (HSIRB) approval was sought, including an application for a waiver for signed consent based on this Code of Federal Regulations (CFO) statement:

the research presents no more than minimal risk of harm to subjects, and involves no procedures, for which written consent is normally required outside of the research context (45 CFR 46.117, 2007, p. 138)

The HSIRB application yielded an “exempt from approval” response, dated March 24, 2008 (see Appendix F). The exemption meant that originally designed communication procedures could be revised and that the usual HSIRB language\(^6^1\) could be removed from the communication, because the study presented no risk of harm to subjects and participation was completely voluntary—so that respondents could terminate participation at any time. Furthermore, both respondents from both groups provided their consented electronically by favorably responding to the recruitment letter.

\(^6^0\) http://pelicanweb.org/solisustv04n04.html#section8wmu

\(^6^1\) This language includes, for example, the following statement: “This consent document has been approved for use on ___/___/2008 for one year by Western Michigan University’s Human Subjects Institutional Review Board. Do not participate in this study before ___/___/2008 or after ___/___/2008.”
The costs to participants in the study were limited to the time involved in reading the checklist and participating in the interview or taking the survey, as well as anticipated fatigue resulting from these activities. Efforts were made to keep the response time and effort as minimal as possible by providing alternative response options and the opportunity to withdraw at any time.

As a thank you for providing feedback on the checklist, participants received a synthesis of the findings from the study and a revised and improved version of the checklist.

Research Procedure

The following section describes the study’s methods of data collection; instrumentation; pretest, pilot test, and think-aloud; the location of data; and data-analysis procedures. Figure 12 displays the timeline for the study.

Figure 12. Study Timeline
Methods of Data Collection

After interested parties responded positively to the requests for participation, a personal e-mail invitation was sent, including the five overarching questions and a copy of the draft SEC.

In the case of the preselected experts, the e-mail also asked about the preferred mode of participation. As indicated earlier, due to the immense time constraints of experts, alternative modes for response included e-mail communication, participation in the Web-based questionnaire, and a telephone interview. Some experts preferred traditional paper-and-pencil communication and requested a copy of the questionnaire in Word (rather than completing the Web-based version). Others participated in an interview and additionally completed the Web-based questionnaire. Again others preferred to simply respond via e-mail.

Listserv members and experts who preferred the online feedback form were pointed to the Web-based questionnaire via a link that was generated and distributed via the Survey Monkey Web system. That is, the checklist was sent via personal e-mail, while the link was sent directly through the survey system. A reminder e-mail was sent on a weekly basis to increase the response rate. The invitation to participate in the survey is provided in Appendix G.

Instrumentation

The interviews and questionnaires were designed as cross-sectional surveys for obtaining evaluative as well as descriptive information about the SEC. However,
and because "qualitative inquiry—strategically, philosophically, and therefore methodologically—aims to minimize the imposition of predetermined responses when gathering data" (Patton, 2002, p. 353), the interview protocol emphasized open-ended questions related to omissions, commissions, and errors in the checklist. The key questions are enumerated in the interview guide (see Appendix H).

As shown, the researcher anticipated the emergence of additional questions during the interviewing process and provided room for the expert to determine the direction of the interview. In essence, the key questions functioned as detail-oriented probes that were followed up with elaboration or clarification probes, such as, "Would you elaborate on that?" (Patton, 2002, pp. 372-374).

Additionally, the interview protocol allowed for adding questions about emerging issues as a result of incoming data that repeatedly emphasized a specific omission, commission, error or problem.

The self-administered questionnaire (see Appendix I) was implemented and administered via Survey Monkey, a Web-based survey software, but also could be requested as a file in Microsoft Word or Rich Text Format. In its Web-based form, the questionnaire consisted of five screens. Screen 1, displayed in Figure 13, provided the general introduction to the study and the survey, identifying the university, principal investigator, and student investigator; communicating appreciation for participation; explaining the tasks involved, anticipated time for required to read the checklist and complete the questionnaire, and modes for disseminating aggregated findings and the revised checklist (Covert, 1984).
Western Michigan University  
Department of: Interdisciplinary Evaluation  
Principal Investigator: Dr. Michael Scriven  
Student Investigator: Daniela Schroeter

Thank you very much for agreeing to provide feedback about the sustainability evaluation checklist that was sent to you via e-mail. You will be asked to respond to a set of questions related to the checklist's content and usefulness. You will also be asked to provide general information about yourself.  

The survey will require approximately 20 minutes of your time in addition to reading the checklist. All responses will be treated confidentially. That means that your name will not appear on any papers on which this information is recorded. I will disseminate aggregated and triangulated findings via my dissertation and potential future presentations and publication.  

To thank you for your time and effort involved in providing this feedback, I will provide you with a summary of key findings from the study as well as the revised and improved checklist.  

If you have any questions or concerns about this study, you may contact me at (001)269-267-8227 or via e-mail Daniela.Schroeter@gmail.com

Thank you very much!

Figure 13. Screen 1 of Web-based Questionnaire

The first screen was followed by two screens with sets of closed-ended and open-ended questions about the SEC. Screen 2 included three sets of rating items to determine the extent to which the checklist would have the characteristics of a valid
and useful framework to guide sustainability evaluation. The rating scale for all three sets of questions were ordinal (0 = not at all, 1 = slightly, 2 = somewhat, 3 = very, 4 = completely). Respondents also had an option to indicate don’t know/understand.

The first question set consisted of questions about the checklist’s characteristics, such as ‘adaptable to differing cultural contexts, comprehensive (i.e., complete), coherent (i.e., items do not contradict each other), concise (i.e., to the point), concrete (i.e., tangible), feasible (i.e., viable), easy to follow, easy to implement, important (i.e., valuable), relevant (i.e., related to the field), useful (i.e., practical, helpful), and valid (i.e., logically correct, legitimate). Respondents could add additional criteria.

The second set of questions inquired about the checklist’s usefulness for considering all aspects of sustainability evaluation, ethics, and those in need; for determining cost-effective ways for evaluating sustainability; for developing appropriate evaluation methodology and evaluating sustainability evaluations; for identifying criteria of specific relevance to sustainability evaluations, defensible information sources, and information needs of evaluation clients; for planning and designing sustainability evaluations; and for promoting evaluation use. Again, respondents could add additional items.

The third set of questions concerned potential impacts of using the checklist and asked if the use of the SEC would (1) change the way in which sustainability evaluation is conducted and (2) improve sustainability evaluation. Respondents could add additional items if desired.
Screen 3 included six open-ended questions about the checklist’s strengths and weaknesses as well as means for improving the checklists:

(1) What is missing from the checklist?

(2) What components or checkpoints are not necessary?

(3) What areas are confusing?

(4) Are there any other errors or problems that need to be addressed?

(5) What, if any, are the strengths of the checklist?

(6) Do you have any suggestions for how to improve the checklist?

Screen 4 included questions about the respondent. Variables of interest to this study included number of years involved in monitoring and/or evaluation (in order to gain insight into respondents’ levels of experience and expertise); roles in monitoring and/or evaluation; regions in which the respondent conducts evaluation\(^{62}\) (to determine the extent to which respondents were involved in development efforts); the home country of the respondent; primary organizational affiliation; sectors in which the respondent conducted evaluation; training in monitoring and evaluation (not only statistics and measurement); type of training; if the respondent considered sustainability when conducting evaluations and if so, how; the highest educational degree of the respondent; areas of expertise; and experience with evaluation checklists.

The final screen (Figure 14) included a thank you note to the respondent as well as contact information to follow-up with the research team.

\(^{62}\) The United Nations’ classification of 21 regions was used.
Measures

Checklist validity was evaluated on two primary dimensions: (A) Accuracy and (B) Utility. Items for each construct were selected purposefully by considering existing metaevaluation checklists (Stufflebeam, 1999c), evaluation standards (Joint Committee on Standards for Educational Evaluation, 1994), and elements of good
checklists as put forth by Stufflebeam (2000) and Scriven (2007). Table 9 displays the items for each scale. Accuracy was measured with 13 items and utility was measured with 12 items.

Table 9
Accuracy and Utility Scales

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The SEC is comprehensive (i.e., complete)</td>
<td>1. The SEC is adaptable to differing cultural contexts.</td>
</tr>
<tr>
<td>2. The SEC is coherent (i.e., checkpoints do not contradict each other)*</td>
<td>2. The SEC is concise (i.e., to the point).</td>
</tr>
<tr>
<td>3. The SEC is valid (i.e., logically correct, legitimate)*</td>
<td>3. The SEC is concrete (i.e., tangible).</td>
</tr>
<tr>
<td>4. The checklist is useful for considering all aspects of sustainability evaluation</td>
<td>4. The SEC is feasible (i.e., viable).</td>
</tr>
<tr>
<td>5. The checklist is useful for considering ethics in sustainability evaluation</td>
<td>5. The SEC is easy to follow.</td>
</tr>
<tr>
<td>6. The checklist is useful for considering those in need.</td>
<td>6. The SEC is easy to implement.</td>
</tr>
<tr>
<td>7. The checklist is useful for developing an appropriate evaluation methodology.</td>
<td>7. The SEC is important (i.e., valuable).</td>
</tr>
<tr>
<td>8. The checklist is useful for evaluating sustainability evaluations.</td>
<td>8. The SEC is relevant (i.e., related to the field).</td>
</tr>
<tr>
<td>9. The checklist is useful for identifying criteria of specific relevance to sustainability evaluations.</td>
<td>9. The SEC is useful (i.e., practical, helpful).</td>
</tr>
<tr>
<td>10. The checklist is useful for identifying defensible information sources.</td>
<td>10. The checklist is useful for determining cost-effective ways for evaluating sustainability.</td>
</tr>
<tr>
<td>11. The checklist is useful for identifying information needs of evaluation clients.</td>
<td>11. The checklist is useful for promoting evaluation use.*</td>
</tr>
<tr>
<td>12. The checklist is useful for planning and designing sustainability evaluations.</td>
<td>12. Using the SEC would change the way in which sustainability evaluation is conducted.*</td>
</tr>
</tbody>
</table>

*Note: Items that are marked with an asterisk were removed from the scales or are discussed separately as a result of item analyses (see Chapter V).
To assess whether the items have been appropriately classified within each scale and whether revisions of the scales were necessary, a multi-scale item analysis was conducted. Both convergent and discriminate validity were determined via the item analyses. Additionally, the reliability of the scales was assessed via coefficient alpha estimates.

**Pretest, Pilot Test, and Think-aloud**

An informal pretest was conducted with two evaluation/development experts to determine whether the questionnaire would be adequate for the anticipated respondents. After feedback from the first expert (who is specifically knowledgeable of the participant respondent group) was obtained, the questionnaire was significantly shortened to better respect respondents' time constraints, clarify questions, and achieve a higher response rate. The revised version was then provided to the second expert whose feedback yielded further changes and refinements regarding specific wording of items.

After these revisions were made, current and former interdisciplinary evaluation doctoral students were invited to participate in a pilot test of the questionnaire. Ten students provided feedback. Many respondents tracked changes throughout the test version of the SEC, which indicated ways the instrument could be improved (e.g., clarifying checkpoints and correcting spelling and grammar) prior to implementing the study with experts and practitioners. Additionally, the pilot test allowed for incorporating further refinements into the questionnaire. For example,
instructions and items were clarified, in some cases rewritten, and the rating scales were reconsidered. The pilot study was re-evaluated in an Evaluation Café presentation (Schröter, 2008b), yielding another set of changes to the questionnaire, including further reduction in the length of the survey, as well as changes in the rating scales.

**Location of Data**

All data were stored in password-protected files on the researcher's personal computer. Data obtained via e-mail were copied into Microsoft Word for purposes of data analysis. Telephone interviews were digitally recorded, saved in electronic files, and selectively transcribed. As a time-saving measure, only information of importance to the investigation was transcribed (Flick, 2006b). All data will be stored in password-protected electronic files in the researcher's home for no less than three years. Raw questionnaire data will be stored in a password-protected database. Table 10 summarizes the modes of communication, resulting data types, and means for storing the data.

**Data-analysis Procedures**

This study used descriptive statistics and inductive data analysis to compress the data, relate it back to the questions of interest, and develop rationales for improving the checklist (Hoepfl, 1997; Patton 2002).

Table 10
Modes of Communication, Data Types, and Data Storage

<table>
<thead>
<tr>
<th>Mode of communication</th>
<th>Type of data</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-mail</td>
<td>Written</td>
<td>Text file</td>
</tr>
<tr>
<td></td>
<td>Electronic text file</td>
<td>Database</td>
</tr>
<tr>
<td></td>
<td>Asynchronous</td>
<td></td>
</tr>
<tr>
<td>Telephone or face-to-face interview</td>
<td>Oral</td>
<td>Audio file</td>
</tr>
<tr>
<td></td>
<td>Electronic voice file</td>
<td>Database</td>
</tr>
<tr>
<td></td>
<td>Synchronous</td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Written</td>
<td>Text file</td>
</tr>
<tr>
<td></td>
<td>Electronic text file</td>
<td>Database</td>
</tr>
<tr>
<td></td>
<td>Asynchronous</td>
<td></td>
</tr>
</tbody>
</table>

First, all qualitative data (including data from the questionnaire and the interviews) were scrutinized for facts describing the checklists strengths and weaknesses as well as for means to improve the too. That is, data analysis was concerned with all information pertaining to omissions, commissions, errors, problems, strengths, and means for improvement).

Second, potential linkages among themes were explored under consideration of respondents’ expertise and the criteria for good evaluation checklists. That is, findings were triangulated based on data from different respondent groups and characteristics (i.e., data triangulation) and different data collection methods (i.e., methodological triangulation) (see Patton, 2002, p. 247). In essence, the ratings from the questionnaire were used to inform the qualitative information gained via open-ended questionnaire items and interviews. Background information about the respondents were used to determine whether their feedback warranted changes in the
checklist in general and whether respondents’ theoretical assumptions could be linked to their opinions about the checklist. The background information also functioned as a crosscheck to illuminate the cultural validity of the checklist.

Third, recommendations for improving the test SEC were considered in terms of their utility for checklist redevelopment and decisions were made about changes to be made.

Chapter V presents the results from the study, including the item analysis and reliability analyses of the measures used within the questionnaire, a description of the respondents, and a summary of the qualitative data analysis. Additionally, Chapter V discusses the implications for checklist improvement and actions taken based on the study’s results.
V. RESULTS

As illustrate in Figure 15, this chapter presents the results from the review of the SEC in five parts.

Results of SEC Validation Study

1. Respondents by Response Medium
   - Questionnaire
   - E-mail
   - Interview

2. SEC Validity Rating Scales
   - Item Analyses
     - Accuracy
     - Utility
     - Other
   - Descriptive Statistics

3. Qualitative Results
   - Cross-Item Analysis
     - Weaknesses
     - Strengths
   - Recommendations
   - Cross-Section Analysis

4. Triangulation

5. Summary and Implications

Figure 15. Chapter V Overview

In the first part of this chapter, the characteristics of the respondents are described under consideration of the response medium (questionnaire, interview,
and/or e-mail). Specific attention is paid to the diversity of perspectives and experiences represented in the sample and assessed within the questionnaire.

The second part of the chapter focuses on the quantitative information collected via validity rating scales. First, findings from multidimensional item analyses (Green & Salkind, 2003) are presented to illuminate the reliability and validity of the “accuracy” and “utility” subscales. These subscales are further explored with descriptive statistics. At the end of this part, additional checklist characteristics identified by the respondents are discussed.

In the third section of the chapter, attention shifts toward the qualitative data provided by survey participants, e-mail respondents, and interviewees. That is, the checklist’s strengths and weaknesses are discussed under specific consideration of omissions, commissions, confusions, errors/problems, strengths, and suggestions for improvement. Using inductive data analysis to compress the data, survey responses were scrutinized via cross-case, cross-item, and cross-sectional analyses to illuminate themes and patterns.

The final part of the chapter presents findings from triangulating the results from the descriptive statistical analysis and the qualitative data analysis. The chapter concludes with a brief summary of key findings and discusses implications for improving the SEC.
Respondents

Participants in the study included evaluation, international development, and sustainable development experts and practitioners. Initially, 44 pre-identified experts were invited to participate in the study through interviews. Twenty of these individuals took part in the study via the questionnaire, informal e-mail, and/or interviews (see Table 11). As shown, many of these experts preferred other means of communication rather than an interview.\(^6\)

Table 11
Pre-identified Respondents

<table>
<thead>
<tr>
<th></th>
<th>Questionnaire</th>
<th>Interview</th>
<th>E-mail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Interview</td>
<td>2</td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>E-mail</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>9</td>
<td>5</td>
<td>20</td>
</tr>
</tbody>
</table>

Listserv invitations and snowball sampling (see Chapter IV) yielded a total of 132 experts and practitioners who (a) indicated an interest in participating in the study, (b) requested and received the SEC, and (c) were provided with a link to the questionnaire or an alternative response format. Of these individuals, 70 percent (91 individuals) completed the questionnaire. Some of these individuals provided

---

\(^6\) While some had scheduled interviews, these could not be implemented as planned due to unforeseen time constraints or emergent obligations of experts, and thus are not presented here. Others (n = 6) agreed to participate in the survey, but did not respond within the timeframe of the study.
additional comments via e-mail, and one of these respondents engaged in a phone
conversation (interview 3 in Appendix J).

Combining the targeted experts \((n = 20)\) with the self-selected respondents \((n = 91)\), yielded a total sample of 111 individuals who participated in the study.

Exploration of differences between invited and self-selected questionnaire
respondents indicated no systematic differences in terms of levels of experience.
Thus, data were not analyzed by type of respondent, but in aggregate only.

**Questionnaire Respondent Characteristics**

Questionnaires were completed via a MS Word document or a link to a Web-based
survey. Responses that were submitted in MS Word were entered into the Web-based
survey system. The Web-based questionnaire was closed on May 11, 2008, and
data were downloaded into a spreadsheet. A total of 106 individuals had accessed the
questionnaire. One respondent was removed, because no data was provided, yielding
105 respondents. The following describes the respondents' background \((N = 105)\).

Background information was not provided consistently, that is six percent of
the respondents did not respond to questions about their backgrounds, and some items
were skipped by up to 10 percent of the respondents. Therefore, the numbers of actual
respondents per question are indicated within the discussion.

*Experience in monitoring and/or evaluation.* Questionnaire respondents \((n =
96)\) conveyed between 1 to 40 years of experience in monitoring and/or evaluation \((M
= 11, SD = 9)\). Together, these individuals brought 1,088 years of experience to bear
on the checklist. While this number does not reflect cumulative knowledge, it provides an idea about the breadth of experience some respondents possess.

*Organizations, functions, and sectors.* The primary organizational affiliation of respondents ($n = 96$) is displayed in Figure 16. As shown, the majority are working in international nonprofit organizations (31%), education (16%), or national nonprofits (15%). Many respondents also indicated affiliation with other types of organizations (15%). These included, for example, free-lance consulting agencies, research organizations, local governments, independent federal agencies, and “none” (i.e., retired professors or professionals).

*Figure 16. Organizational Affiliations*
Respondents (n = 98) also indicated a variety of roles in monitoring and evaluation. Multiple responses were possible and results indicate that most individuals have more than one function within their organizations. Most frequently, respondents indicated the following roles: external evaluator (63%), internal evaluator (61%), program manager (42%), and/or evaluation funder (10%). Several respondents (22%) also reported “other” roles, including:

- critical friend
- advisor
- administrator
- DM&E manager
- evaluation educator/instructor/trainer
- evaluation researcher
- evaluation consultant
- metaevaluator
- theorist
- program quality specialist
- project coordinator
- retired professor
- evaluation contractor
- sustainability researcher

The variety of organizational affiliations and roles within organizations was also evident in terms of respondents’ (n = 94) sectorial involvement. As shown in Figure 17\textsuperscript{65}, most respondents conduct evaluation in the context of social services (56%) or education (55%), though many (54%) indicated “other sectors” that were

\textsuperscript{65} Multiple responses were possible.
not specified within the questionnaire, including health (e.g., care, nursing, HIV/AIDS, nutrition, etc.), environment, conflict resolution and peace building, governance, communications, religious services, research, and social and economic awareness, and others.

![Sectors Diagram]

**Figure 17. Sectors**

*Work regions and home countries.* Some individuals \((n = 46)\) reported the *regions* in which they were working. Although the response rate to this question was relatively low, respondents represented most regions\(^{66}\) around the world. The most frequently reported regions were Eastern Africa, North America, South-Eastern Asia,

---

\(^{66}\) The categorization was based on the United Nations Classification available at http://www.un.org/depts/dhl/maplib/worldregions.htm
Western Africa, Western Europe, South America, and Eastern Europe (see Figure 18).

Figure 18. Regions Where Respondents Work

This diversity is also evident in the origins of the respondents (n = 94), as indicated by their home country (see Figure 19). As shown, the largest number of respondents were North Americans (from the United States and Canada), followed by Europeans (from Austria, Italy, Bosnia and Herzegovina, Denmark, the United Kingdom, Ireland, Ukraine, the Netherlands, Spain, Switzerland, and Germany), Asians (from Japan, Bangladesh, Iran, Pakistan, Tajikistan, India, Nepal, Malaysia,  

67 Multiple responses were possible.
Philippines, Vietnam, Indonesia, and the Republic of Georgia), Africans (from Tanzania, Uganda, Zambia, Ethiopia, Kenya, Mali, Nigeria, Senegal, Sierra Leone, Ghana, and Cameroon), Latin Americans (from Argentina, El Salvador, Guadeloupe, French West Indies, and Trinidad and Tobago), and individuals from Oceana (namely from Australia and New Zealand).  

Figure 19. Respondents' Home Continents

*Educational background.* Respondents were also asked about their educational background in general and specifically in terms of formal training in monitoring and/or evaluation. As shown in Figure 20, the majority of respondents ($n = 96$) obtained graduate (49%) or postgraduate (44%) degrees. Only seven percent indicated that they had no graduate degree, but did have undergraduate degrees (7%). No respondent indicated less than an undergraduate degree.

---

68 Again, the categorization was based on the United Nations Classification available at http://www.un.org/depts/dhl/maplib/worldregions.htm
Additionally, many respondents (78%) indicated that they had received formal training in monitoring and/or evaluation (based on \( n = 98 \)). Ninety-five of these respondents indicated that this training was obtained on the job (75%), in university programs (45%), or off the job (24%). Fourteen percent indicated additional outlets for receiving such training, for example, via other area-specific training workshops (e.g., project management, auditing) or self-study.

![Figure 20. Respondents’ Education Level](image)

Consideration of sustainability. Of special interest to this study was whether respondents consider sustainability in their evaluations. The vast majority (80%) indicated that sustainability was usually considered (based on \( n = 95 \)). Seventy-four respondents provided descriptions of how they considered sustainability. As
illustrated in the selected examples below, the respondents have varied perspectives and favor different approaches to sustainability evaluation:

The following is an excerpt of the standard we use to assess the quality of our evaluations' assessment of sustainability: "Sustainability is the durability of programme results after the termination of technical cooperation. Static sustainability is the continuous flow of the same benefits, set in motion by the completed programme, to the same target groups. Dynamic sustainability is the use or adaptation of programme results to a different context or changing environment by the original target groups and/or other groups. In the case of formative evaluations that take place mid-way through the intervention process, sustainability may not be an applicable criterion, unless the intervention has been going on for a number of years. Even in evaluations carried out towards the end of interventions, evaluators may have to rely on plausible assumptions about likely sustainability, as it can take several years before it is clear whether interventions are sustainable or not. Ex-post evaluations, three to five years after final project/programme funding has ended, would allow evaluators to determine whether adequate attention was paid to questions of sustainability. Such an evaluation would also allow for an assessment of impact. To meet this standard the evaluation needs to cover: § Assessment of the extent to which the programme/project results have had or are likely to have lasting results after programme/project termination and the withdrawal of external resources. Evaluators should assess whether the planning process covered sustainability questions adequately, and includes realistic strategies for UNFPA withdrawal (i.e. phasing out over an agreed period of time). § Assessment of the factors affecting sustainability on the basis of the priority that stakeholders assigned to the programme/project, this includes their readiness to continue supporting or carrying out specific activities; replicate the activities in other regions or sectors of the country, or to adapt programme/project results in

Excerpts present responses from different individuals.

Examples were edited and shortened for the purpose of the dissertation.
other contexts. For example, the evaluation should assess the likelihood of governments or other counterparts continuing to support the initiative after UNFPA withdraws. Assessment of sustainability should also include analysis of the extent to which local capacity has been supported and developed to take over programming.

It is always a criterion demanded by donors in the DAC meaning. I usually avoid using classical economic indicators (because the nature of projects, which almost all are micro, only we consider some of them in productive applications), and I do a revision of the identified outcomes from different dimensions, trying to do a prospective valuation. These dimensions are: environmental, gender, technology, institutional issues, political issues, participation, social-cultural.

In ex post evaluation, by assessing the extent of (1) presence of the intervention and how it has changed since it was originally introduced; (2) likelihood that it will continue based on an assessment of the organizational supports and perspectives of key decision making informants.

If an intervention is still in process, by assessing (1) impact to date (2) perceptions of the value of the impacts (c) cost of continuation and feasibility of having sufficient material resources to continue (4) organizational support—material and psychological—for its continuation.

Prospectively, by looking at type of sustainability sought (of activities, results, capacity, ideas), strategies undertaken to achieve this, and existence of predictors of sustained activities (e.g., engagement of champions, evidence of effectiveness, etc.). Retrospectively, by following up short-term projects after completion to see what is left 6-12 months after funding has ended. I usually evaluate sustainability as a criterion of effectiveness (e.g., DAC/OECD criteria) emphasizing such elements as:
• Degree of local participation and consultation in identification of project goals, outcomes, etc.

• Extent to which local processes and structures are used (e.g., existing governance mechanisms and systems, compatibility with local beliefs and cultural practices)

• Extent to which local resources (human, financial and natural/physical) are utilised.

Evaluation of Sustainability—i.e., to what extent the primary beneficiaries of the programme will be able to continue with and expand upon the positive outcomes and manage the negative consequences of a programme, once the technical and financial support has been removed. I prefer a return to a project area twelve months after cessation to assess if the overall impact has been retained and what can be done to reverse any slippage.

I work with clients to ensure that they place sustainability questions throughout the life cycle of a program, e.g., are they diversifying funding, are they engaging with partners who will be there when the initial funding ends, are they considering the potential for changing environment/context/clients in the future, etc.

Primarily through post-project/program visits to key stakeholders to assess if, and to what degree, the positive intended goals, objectives and both intended and unintended results are continuing via local community support, broader-level public support (district/region) and nationally (policy).

We combine a network perspective with a systems approach, using outcome mapping and other tools, side-stepping the attribution gap. The dimensions of sustainability are incorporated into the relationships, e.g., as learning across frames manifested observable behaviour change. Systems diagramming helps establish cause and effect relationships. We like the approaches of Rick Davies and Bob Williams.
It depends upon the type of evaluand you are dealing with; however, in case of development projects, the sustainability measures should be the long-term dynamics of the project activities and outcomes. The projects should not stop with the financial discontinuation of the donors. There should be mechanisms in such a way that the project activities and outcomes would be continued in benefiting the project beneficiaries.

Sustainability subject matter is usually addressed by project/program design. Preliminary info about sustainability, communities’ capacity for sustainability, local partners’ and other stakeholders’ capacity and vulnerability is collected during the assessment. During the baseline measurement, we establish benchmarks for sustainability (organizational, financial, social, economic . . . depending on nature of programmatic intervention) that are indicators on programme outcome level and integrated projects goal level and depending on type of evaluation, we measure progress towards achievements. Confusing part is usually measurement of contribution and attribution towards the achievement of sustainability objectives. Basically sustainability needs to be addressed from the very beginning—assessment needs to be properly followed by functional M&E system. During the design it is very important that sustainability and sustainability measures are formulated as seen by major recipients/stakeholders in order not to diminish existing coping mechanisms by our programmatic intervention. Most times, projects do not have an exit strategy in place and as soon as they leave, the participants go back to the same old situation. So, I normally look at strategies put in place to keep the participants doing what the project sought to improve. Again, projects that are planned top-down usually collapse after the project. So I look out for processes of selecting a particular intervention. If it’s people centered, then there’s a likelihood that the beneficiaries will continue to improve on the intervention. Again, if local structures and materials as well as indigenous knowledge are used, there is likelihood that the project can be sustained.
I have a template checklist that will improve as a result of this exercise. I tailor the checklist for every evaluation in consultation with the evaluand. If we come to agreement, I do the evaluation (for a handsome amount of course), but if I see that the evaluand is playing games, I simply walk away. In the case of ISO audits, I go by the book—I try to avoid submitting severity 1 findings, but generally submit severity 2 and 3 findings that require corrective action. When there is a severity 1 finding, I am generally successful in having it partially corrected on the spot, so I can submit as severity 2. I have never pulled a certificate.

How the existing (good ones) models can be taken forward after the project completion within the existing project area or in new area. We assess whether this is relevant to the communities, providing positive impacts, are more efficient than other model used in the project area, are the partners interested to integrate into their programme, is there capacity and incentive (economic and social) of local beneficiaries, is there local mechanisms (groups/revolving funds, etc.) to sustain the interventions, etc.

Checklist experience. Finally, respondents were asked whether they had experience working with evaluation checklists. Based on 97 respondents, 64 percent had worked with evaluation checklists, including, for example, checklists available on WMU’s EC Checklists Web site (e.g., Scriven’s Key Evaluation Checklist, Patton’s Utilization-focused Evaluation Checklist, the reporting checklist, and others by Stufflebeam and Scriven), personally developed checklists, checklists that were developed “in-house,” and available checklists from various organizations (e.g., European Commission Evaluation Standards, OECD DAC checklists, WHO health systems, project cycle management basic framework, checklists of the Dutch Ministry
Characteristics of Interviewees and E-mail Correspondents

Interviews were conducted with 10 individuals, seven of whom also completed the questionnaire and two of whom also provided some comments via e-mail. E-mail responses were provided by five individuals, two of whom also responded via the questionnaire. The remaining three individuals responded by e-mail only. All but one of these respondents belonged to the pre-identified expert group and all, including the follow-up interview, brought extensive experience and scholarship in at least two of the three categories: evaluation, sustainability, and/or international development. Eight of these respondents were from the United States, eight from Europe, two from Oceana, and one from Canada. These individuals work in academe; international development agencies; nonprofit, for-profit, and/or multilateral organizations; and as independent consultants.

In sum, the respondents included experts and practitioners with extensive experience related to monitoring and evaluation in national and international development contexts. Affiliations, roles, and sectorial involvement varied tremendously, as did the regions around the world where respondents work and live. In addition to extensive experience, respondents revealed high levels of education, formal and evaluation-specific. Furthermore, most had experience with sustainability evaluations and many had worked with checklists.
SEC Validity Rating Scales

The questionnaire included three sets of items intended to assess the SEC’s validity, measure via accuracy and utility items (see Chapter IV). The 25 rating items were recoded for analysis in SPSS from 0 = not at all to 4 = completely. "Don’t know/understand responses were recoded as missing data. Missing data on the rating items were analyzed, indicating that 0 percent to 18 percent of the data per item were missing. On average, less than 6 percent of the data were missing for a given item. Correspondingly, missing data were substituted using mean replacement for the item. Comparisons of item means before and after the substitution indicated inconsequential differences ranging from 0.09 to -0.03 on the five-point scale. Across all items, these mean differences were less than .005, suggesting that the mean substitution procedure did not alter the results to any noticeable extent.

Reliability and Validity of the Accuracy and Utility Subscales

Item analyses were conducted on the 25 items hypothesized to assess checklist validity via accuracy and utility scales. Several items were more strongly correlated with the other scale than their own scale, including these:

---

71 The reason for that was bi-fold: First, “don’t know/understand” is a semantically unsound stem and the response does not allow for differentiating between who did not know and who did not understand. Second, the item was initially included to allow multilingual respondents to opt out from responding to a question. However, as missing data were present in the response set and the true meaning of the items could not be captured, these responses were coded as missing.

72 A measurement model could not be constructed due to the limited amount of data. Additionally, principal component analysis and confirmatory factor analysis were deemed inappropriate due to the
A. Change the way in which sustainability evaluations is conducted

B. Promoting evaluation use

C. Determining cost-effective ways for evaluating sustainability

D. Coherent

E. Valid

F. Developing an appropriate evaluation methodology

G. Improve sustainability evaluation

Based on these results, careful consideration, and an additional item analysis, the two scales were refined and the items were dropped from the scales.  

To assess the convergent and discriminant validity of the SEC validity scales, each item was again correlated with its own scale (without the removed items) and with the other scale. The results of the analysis are displayed in Table 12.

In support of validity of the measures, items were more correlated with their own scale (convergent validity) than with the other scale (discriminant validity). Coefficient alpha was computed to obtain internal consistency estimates of lower-bound reliability for the two scales. The alphas for the Accuracy and Utility subscales were .87 and .89, respectively.

73 In essence, Item A is semantically confusing, as the direction of the change was not indicated. Therefore, the item is not further discussed in this chapter. Items B to G could not be discriminated for the hypothesized scales and were dropped from further analysis.
Checklist Accuracy

As shown in Table 12, the checklist’s accuracy was assessed via nine items. To assess the extent to which the SEC was perceived as accurate, the mean total scale score was computed and divided by the maximum possible total scale score. Results indicate that checklist accuracy was perceived as 67 percent, which corresponds to “very accurate” on a five-point scale.

Table 12

Correlations of Each SEC Validity Items With its Own Scale and the Other Scale (Accuracy or Utility) After Removing Inadequate Items

<table>
<thead>
<tr>
<th>Items</th>
<th>Accuracy</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive (i.e., complete)</td>
<td>0.52</td>
<td>0.38</td>
</tr>
<tr>
<td>Considering all aspects of sustainability evaluation</td>
<td>0.68</td>
<td>0.39</td>
</tr>
<tr>
<td>Considering ethics in sustainability evaluation</td>
<td>0.59</td>
<td>0.32</td>
</tr>
<tr>
<td>Considering those is need</td>
<td>0.47</td>
<td>0.45</td>
</tr>
<tr>
<td>Evaluating sustainability evaluations</td>
<td>0.70</td>
<td>0.34</td>
</tr>
<tr>
<td>Identifying criteria of specific relevance to sustainability evaluations</td>
<td>0.61</td>
<td>0.36</td>
</tr>
<tr>
<td>Identifying defensible information sources</td>
<td>0.59</td>
<td>0.43</td>
</tr>
<tr>
<td>Identifying information needs of evaluation clients</td>
<td>0.60</td>
<td>0.40</td>
</tr>
<tr>
<td>Planning and designing sustainability evaluations</td>
<td>0.67</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Table 12—continued

<table>
<thead>
<tr>
<th>Utility Items</th>
<th>Accuracy</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptable to differing cultural contexts</td>
<td>0.38</td>
<td>0.46</td>
</tr>
<tr>
<td>Concise (i.e., to the point)</td>
<td>0.29</td>
<td>0.57</td>
</tr>
<tr>
<td>Concrete (i.e., tangible)</td>
<td>0.56</td>
<td>0.76</td>
</tr>
<tr>
<td>Feasible (i.e., viable)</td>
<td>0.29</td>
<td>0.68</td>
</tr>
<tr>
<td>Easy to follow</td>
<td>0.32</td>
<td>0.63</td>
</tr>
<tr>
<td>Easy to implement</td>
<td>0.44</td>
<td>0.77</td>
</tr>
<tr>
<td>Important (i.e., valuable)</td>
<td>0.42</td>
<td>0.48</td>
</tr>
<tr>
<td>Relevant (i.e., related to the field)</td>
<td>0.41</td>
<td>0.65</td>
</tr>
<tr>
<td>Useful (i.e., practical, helpful)</td>
<td>0.65</td>
<td>0.79</td>
</tr>
</tbody>
</table>

To determine the checklist’s strengths and weaknesses in terms of its performance on the accuracy items, the analysis was repeated for each item (see Figure 21). The percentages for the items ranged from 59 percent for considering ethics in sustainability evaluation to 73 percent for planning and designing sustainability evaluations, corresponding to a judgment of somewhat to very accurate.

As shown, the checklist performed least well in terms of considering ethics (59%), identifying defensible information sources (60%), and considering those in need (60%). Checklist accuracy was perceived highest in terms of planning and designing evaluations (73%), comprehensiveness (73%), and considering all aspects of sustainability evaluation (72%).
Planning and designing sustainability evaluations
Comprehensive (i.e., complete)
Considering all aspects of sustainability evaluation
Identifying criteria of specific relevance
Evaluating sustainability evaluations
Total Accuracy
Identifying information needs of evaluation clients
Considering those is need
Identifying defensible information sources
Considering ethics in sustainability evaluation

Perceived Percentage of SEC Accuracy

Figure 21. Checklist Performance on Accuracy Subscale

The percentages convert to the following rubric: 0% to 20% = Not at all accurate, 20% to 40% = Slightly accurate, 40% to 60% = Somewhat accurate, 60% to 80% = Very accurate, and 80% to 100% = Completely accurate.

Figure 22. Frequencies Distribution of Checklist Performance on Accuracy Items
As illustrated in Figure 22, only 4 to 13 percent thought the characteristics were not at all or slightly apparent. Furthermore, it is notable that 40 to 76 percent of respondents perceived the characteristics as very or completely apparent. Even in the lowest performing item, 40 percent of respondents indicated that considering ethics in sustainability evaluation was very much or completely apparent in the SEC, while 12 percent said that ethics was considered slightly or not at all.

**Checklist Utility**

The items assessing checklist utility are displayed in Table 12. To assess the extent to which the SEC was perceived as useful, the mean total scale score was computed and divided by the maximum possible total score scale. Results indicated that checklist utility was perceived as 67 percent, which corresponds to “very useful” on the five-point scale. To determine the checklist’s strengths and weaknesses in terms of its performance on the utility items, the analysis was repeated for each item (see Figure 23).

As shown in Figure 23, the checklist received the lowest ratings for easy to implement (52%), easy to follow (61%), and concise (62%). Checklist utility was perceived highest in terms of its importance (82%), relevance (79%), and usefulness (74%).

As illustrated in Figure 24, the lowest performing items received more frequently lower ratings than higher performing items. However, less than 10 percent of the respondents indicated that utility characteristics were not at all apparent in the checklist. Conversely, it is notable that more respondents perceived the characteristics as very or completely apparent (between 30% and 90%) than as not
at all or slightly apparent (between 2% and 23%). That is, 30 percent of the respondents said that the SEC would be very much or completely easy to implement, while 23 percent perceived it as slightly or not at all easy to implement.

![Checklist Performance on Utility Subscale](image)

**Figure 23. Checklist Performance on Utility Subscale**

**Other Checklist Characteristics**

In addition to the items discussed above, respondents had an opportunity to list and rate items pertaining to (a) characteristics of the SEC, (b) potential uses, and (c) consequences of use. Emergent themes, striking issues, and recommendations are discussed by the three aforementioned categories. Most respondents chose to provide additional items without providing ratings.

---

75 The percentages convert to the following rubric: 0% to 20% = Not at all useful, 20% to 40% = Slightly useful, 40% to 60% = Somewhat useful, 60% to 80% = Very useful, and 80% to 100% = Completely useful.
1. Characteristics of the SEC. A few respondents indicated that more attention is needed in terms of data sources, collection, and analysis as well as related concerns. One respondent commented on Section 3 of the SEC:

*It totally skips the step of data analysis—ignored both under the "collect data" bullet and the "synthesis" bullet. There is one good, often overlooked, dimension of sustainability evaluation which is highlighted here that I've not seen emphasized elsewhere—the time and space dimensions of sustainability, including timing of costs and timing of benefits. This is good enough to deserve more emphasis.*

Additional responses suggested that the checklist is too academic, thus should be revised to be more comprehensible, easier to understand, and more useful for development practitioners and evaluators. One respondent noted that the SEC “seems to not be clear about whether it is sustainability, likelihood of being sustained, value
of being sustained, etc.,” while another suggested that the SEC “should be useful for both attaining sustainability and maintaining/improving sustainability.”

Furthermore, ethical issues, appropriateness of technology, and the role of strengths and assets are adequately addressed yet, according to respondents. Additionally, the concept of sustainability should be better integrated in the first part of the checklist and distinctions “between sustainability of the evaluand (as an organization) and its program (which relates more to activities and/or impacts)” have to be better specified.

Moreover, one respondent suggested that the checklist would be constrained by the “intellectual framework . . . (e.g., the foregrounding of merit, value and significance). There is an implicit model of what constitutes evaluation that may not be appropriate when dealing with something as ambiguous as ‘sustainability.’”

On the more positive side, respondents conveyed that the checklist was grounded in theory and had conceptual clarity. It would also be very “important as ‘raw material’ to be adapted to various concrete cases.”

2. Potential uses. Respondents reported that the SEC would be somewhat useful for evaluation training and to respond to different demands and concerns of stakeholders. The checklist would be very helpful for beginners, for generating discussion around what is meant by ”sustainability,” for helping to think globally about important aspects of program evaluation, and for identifying all relevant criteria. One respondent said that the SEC is completely useful for “Assisting evaluators in considering the multiple dimensions of sustainability. Evaluators may
select certain sections as needed.” Another suggested that the SEC is a good “starting scenario, to be improved in light of experience.”

3. Consequences of use. Respondents stated that using the SEC would have these positive consequences:

- Identify shortcomings in sustainability evaluation.
- Clarify definitions of sustainability evaluation.
- Raise suggestions for considering evaluation from different point of views.
- Lend more credence to the need for sustainability evaluations.
- Help to think of aspects that one may not have considered.
- Make sustainability evaluation more professional.
- Start to accumulate experience leading to constant improvement.
- Allow the evaluation activity to cover all the necessary elements.
- Allow for comparisons among different evaluations.
- Get an evaluator or program leader to consider more than a few issues around sustainability.
- Assist with project/program design.
- Support multidimensional and trend analysis.
- Help to structure evaluations.
- Lead to adding data-gathering formats/tools to the SEC.
- Greatly enhance evaluation processes as a detour from other conventional methods.

Respondents also stated that using the SEC would have these negative consequences:

- Make sustainability evaluations a very heavy process which is not cost-effective.
- Risk further confusing what it is that people are trying to do.
Strengths, Weaknesses, and Recommendations

All respondents were asked a set of six overarching questions about the SEC that related to omissions (What is missing?), commissions (What components or checkpoints are not necessary?), confusions (What areas are confusing?), errors and problems (Are there any other errors or problems that need to be addressed?), strengths (What, if any, are the strengths of the checklist?), and suggestions (How can the checklist be improved?).

As illustrated in Figure 25, these questions were first screened case by case and coded across cases (cross-case analysis). Second, responses were considered across items (cross-item analysis). Third, responses were analyzed by section (i.e., weaknesses, strengths, and suggestions; cross-section analysis).

The following sections present the findings from each analysis. Note that “item” is used to refer to the overarching categories: omissions, commissions, confusions, errors/problems, strengths, and suggestions. “Code” is used to refer to emerging patterns from responses. “Theme” is applied to groups of codes.

Cross-case Analysis

Each response was coded by applying emergent themes. The coding was iterated five times\textsuperscript{76} to refine the coding scheme and elucidate key themes. The themes were then explored across cases for each item. First, the codes were screened

\textsuperscript{76} This means that text was examined five times by the author of the dissertation. Other coders were unfortunately not available for this project.
for consistency and revised if different terms were used to express the same meaning or if they could be grouped under a more general theme. Second, themes were analyzed across cases within items.

Figure 25. Overview of Qualitative Analysis

Table 13 displays the items, corresponding questions, total number of questionnaire responses for a given item, number of codes applied across responses, and number of unique themes within each item. Themes across responses within items will be further illuminated in the following sections.
Table 13
Overview of Items, Questions, Responses, Codes, and Themes

<table>
<thead>
<tr>
<th>Item</th>
<th>Question</th>
<th>Number of Responses</th>
<th>Number of codes across responses</th>
<th>Number of unique themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omissions</td>
<td>What is missing?</td>
<td>85</td>
<td>85</td>
<td>13</td>
</tr>
<tr>
<td>Commissions</td>
<td>What components or checkpoints are not necessary?</td>
<td>58</td>
<td>52</td>
<td>9</td>
</tr>
<tr>
<td>Confusions</td>
<td>What areas are confusing?</td>
<td>71</td>
<td>68</td>
<td>8</td>
</tr>
<tr>
<td>Errors/problems</td>
<td>Are there any other errors or problems that need to be addressed?</td>
<td>55</td>
<td>50</td>
<td>11</td>
</tr>
<tr>
<td>Strengths</td>
<td>What, if any, are the strengths of the checklist?</td>
<td>79</td>
<td>105</td>
<td>12</td>
</tr>
<tr>
<td>Suggestions</td>
<td>How can the checklist be improved?</td>
<td>75</td>
<td>94</td>
<td>10</td>
</tr>
</tbody>
</table>

Omissions. As indicated in Figure 26, responses within omissions most frequently related to “specifications” not found within the SEC. These specifications were usually highly explicit and included responses requesting more detail related to concepts and terms or the addition of a line-item under an existing checkpoint.

Specification problems are exemplified by the following statements:

*Horizontal Evaluation—In case of direction of evaluation, here may be a tier of Horizontal Evaluation. Generated from and conducted by the participants of a certain activity.*

---

Note that this number does not mean that each response received one code. Some responses were extensive and required multiple codes, while others did not include any information relevant for coding (e.g., see suggestions for improvement).
Page 7—Use of evidence from monitoring and evaluation: You should include sharing of the results of the M&E with the beneficiaries.

At section 2, regarding the evaluand, something should be said about the funding source, how funding is provided and its longevity (e.g. international organizations, governments, foundations, five or more years, one year renewable, matching funds, etc). Also, regarding the direction of the evaluation, it is not that simple (bottom-up or top-down); many times it is a mix, with the funders requiring the evaluation but the implementers very supportive and working together with the evaluator. Others, the implementers are totally hostile to the evaluation and will pose serious limits to the process. Evaluation is generally a messy process.

Figure 26. Omissions: Frequency of Emergent Themes
A second frequently occurring theme within omissions was “tensions” (11%) between simplicity and complexity, necessity and sufficiency, and/or brevity and length. For example, statements reflecting such tensions included the following:

I think the checklist is very much complete. Even "too" complete for me. That leads me to a question: "Will there be a room for improvisation"? Is there a room for participatory process when someone needs to adjust a little bit in order to meet their needs?

First (and second) impression: Nothing relevant is missing. On the contrary, there's a whole lot of information in this checklist, probably more than could be handled in any evaluation.

This is an amazingly thorough project. I did not see anything missing. The "somewhat" check marks in the first scale are only because I found it an overwhelming amount of detail. But excellent points, nonetheless.

The addition of “new (sub) categories” was another theme that occurred commonly. Most often participants referred to the omission of a dimension on politics or issues related to inclusion, exclusion, and participation. Below are some examples of this type of feedback:

There is a lack of discussion of issues and tensions related to exclusion—in terms of both impacts on sustainability (lack of access to education, health and other services creating structural limits on human and social capital; exclusion as noted above from resources) and the evaluation process and the requirement to consciously and proactively seek out excluded groups if they are to be considered at all.
Not enough emphasis on spoilers, i.e. those who would prevent the “success” of the evaluand.

Section 4: My favorite and probably the most helpful. Power is missing! Power relations have a huge influence on sustainability. Something about bureaucratic bottlenecks should be added under organizational characteristics.

Furthermore, “supplementary materials” emerged as a theme. These included requests for inclusion of examples, cases, summary tables, and “quick guides.” For example, one respondent suggested this list of additional information that would be beneficial in the SEC:

- Checklist templates—example matrices, indicators, decision-making systems, e.g. multicriteria analysis, CBA, EIA, health impact assessment, gender impact assessment
- Case study examples
- Legal requirements, e.g. environmental liability, labor standards, human rights
- Bibliography/literature review of useful case studies/methodologies

Other responses indicated a need for “clarification,” made evident by respondents’ misunderstandings of statements in the checklist. Similarly, a lack of “clarity” emerged and was expressed in terms of “too technical” or need to “simplify language.”
Figure 26 also shows that some individuals' responses within omissions emerged as "strengths." These respondents suggested that the SEC is complete and comprehensive, for example:

*It's comprehensive. From my point of view, this is the main strength that a checklist should have, and this checklist has it.*

Finally, there are a few themes that occurred less frequently. These will not be elaborated on in detail here, but suggestions for revising the checklist. For example, "glossary" related to terminology that should be added and "application" refers to the need to widely apply the checklist to enable learning from experience and continuous improvement.

The interviewees and e-mail respondents also commented on omissions and named specifications as well as clarifications as problematic in some respects. One e-mail respondent maintained that "relevance of the sustainability criterion" and "an assessment of financial sustainability" were not sufficiently emphasized. Moreover, the need to stress the political dimension more specifically was communicated.

*Commissions.* As indicated in Figure 27, many respondents suggested that commissions were not a problem in the SEC. This is supported by the second most frequently occurring theme: "Selectivity." Selectivity means that respondents indicated that, while not all checkpoints might be of use in all cases, they should be treated as modules from which users can choose and decide what is important within a given evaluation.
Figure 27. Commissions: Frequency of Emergent Themes

Responses exemplifying “selectivity” are listed below:

I would rather not delete checkpoints. Rather, leave it to the user to decide in any given evaluation which details would be necessary.

I suggest you keep it as comprehensive as possible. Evaluators may select the relevant sections and omit the ones that are not useful for the particular evaluation.

As I understand the checklist to provide examples of what the individual evaluation might contain, I would not omit any of the checkpoints.

I found all of the sections of interest and worth keeping. Frankly, given the paucity of genuine thinking about sustainability in my field, I would encourage breadth of inquiry. Some sections will be more relevant than others, depending on the nature of the system being evaluated.
“Clarity” was another issue that frequently occurred under commissions. Similarly to the lack of “clarity” expressed under omissions, typical responses in this category express a need to simplify language and technicalities. Example responses are displayed below:

- I cannot name specific components which are not necessary, but as a whole, I think it might be too much detailed or dense for a checklist.

- A lot of rhetoric about "evaluand"—not typically used in this way. Too academically oriented. Needs to be a simpler presentation so it is not daunting to users.

- I do not believe there are unnecessary elements, but there must be ways of sharpening and refining the document.

Another reoccurring theme was “reconsiderations,” which relates to aspects that should be rethought when revising the SEC. Here are two examples:

- Page 9: Facets of Cost. The Global Cost might be difficult to count for smaller projects, especially if they have to count the environmental cost (for example).

- Waste reduction (page 7, 9) is a popular aspect of sustainability. I am not sure that it is a critical success factor for most evaluations.

“Part A” of the SEC was perceived as unnecessary by some individuals, which is expressed in the following examples:

- If I read the introduction to the sections correctly, Section A is not about evaluating sustainability or for sustainability, but is a checklist of all the
items that should be considered in general for any evaluation. While doing
the general evaluation well contributes to sustainability, I didn't feel this part
was as important to your question about evaluating for sustainability or
evaluation contributing to sustainability. The completeness is extremely
valuable, but I'm not sure why it's included if the project is about the
sustainability question. Still, it's a lot of really good stuff!

I am not sure if Part A is necessary for this checklist. It somehow distracts
attention from the issue of sustainability. There are other evaluation
checklists such as KEC which guide evaluation efforts in general. This
document does not need to repeat such checklists; it should have a primary
focus on sustainability.

A more radical suggestion would be to cut Sections 1 through 3, and
concentrate on criteria for sustainability, with a framing section explaining
the checklist’s underlying concept of sustainability (distinguishing between
the different levels of institution, program and results) and discussing the
interaction between the main dimensions. To be honest, as an experienced
evaluator, these sections add very little for me.

Additional comments suggested the removal of specific “sections.” One
individual thought that it would not be necessary to discuss general procedures for
evaluation, while others pointed out that the section on criteria of value would not be
essential in all cases:

Section on cost would be less applicable in healthcare in Canada, because
we are publically funded, so generally our projects are funded by granting
agencies and budgets are outlined for a given period of time.
“Overlaps” were another emergent theme within commissions. A response example in this theme group is presented below:

Some questions appear redundant/repetitious: Section 1, subsection 4 (audience) —Q3 intended users is just restating the previous two questions. Section 3, subsection 3 (collect data) — points 1 and 2 repeat point 1 of the previous subsection. Section 3, subsection 3 (collect data) — point 6 repeats points 3 and 4 of the previous subsection. Section 3, sub-section 4 (synthesis) — point 2 repeats point 2 of subsection 2 (set standards).

Finally, comments related to themes such as “tensions” (which already occurred within omissions) and “food for thought” were made less frequently, but were maintained for further analysis and potential checklist improvement.

Confusions. Figure 28 illustrates the emergent themes under confusions. As shown, confusions related to detail/density, food for thought, language, methods, structure, and tensions, whereby some respondents, such as the one quoted below, also indicated that confusion was not a problem:

In general is very clear, perhaps one would need to put it in practice before being sure what is or not confusing. Based only on the checklist, I liked it very much.
The methodology is quite clear and very well expressed throughout the checklist

The most frequently emerging confusions related to “concepts,” that is, evaluation-specific concepts or sustainability-specific concepts. Comments of this sort are presented below:
I think fitting the questions under one of three headings (merit, value and significance) confused me, since some don't fit easily into a single category—the allocation seems arbitrary at times. There may be a more natural ordering of these questions. I've run into problems like this before with checklists. You design a framework (e.g. merit, worth, significance) which you use to understand and gather ideas, but that framework may not be the best framework by which to group and communicate those ideas to other people. Similarly the distinction between "process" and "outcome" doesn't really do it for me. It's too linear and crude a distinction.

The idea of considering both evaluations OF and FOR sustainability is very interesting, but a little bit confusing itself. Maybe "eval FOR" is considered in this checklist with an excessive weight in the idea of "continue positive effects of the intervention." I miss in this point more synergy with environment of the evaluand (not environmental, I mean the context, mainly in local development terms).

![Figure 28. Confusions: Frequency of Emergent Themes](image-url)
“Language” appears to be an issue for potential users of the checklist as well. Comments related to “language” were specifically targeted at the academic and technical nature of the SEC, and include, for example, these:

Depending on how sophisticated the respondent is, the terms might be somewhat confusing, even though they are defined. I imagine that one would need a background in evaluation/research in order to comprehend the terms and phrases and stream of logic. Who is it expected that would be using the tool? If the persons have some exposure and background to evaluation, it should be fine.

Use of terms like mnemonic device when memory aid is sufficient: Are you writing to impress your academic peers or for utilization in the field?

Related to language, as well as to the discussion on commissions, is the theme “detail/density,” which refers to remarks that the SEC is too complex and thick, even for the distinguished and highly educated respondent group. Examples of responses within this theme are listed below:

It DOES take some concentration.

The criteria for sustainability seem somewhat managerial at first sight. A lot of complex issues get squeezed into little space. That's probably the nature of a checklist, though.

Furthermore, comments related to “methods” frequently occurred. Responses in this category usually related to means for implementing the SEC or to questions
about how to best approach certain issues within the SEC. Some examples are provided below:

Section 1 could be managed by the evaluator after a nonstructured talk with the evaluation client. This section probably appears more complex than it is.

When it comes to setting standards (page 4) are these standards set by the evaluation team? And if so, does this team consult with the beneficiaries (stakeholders/impactees)? I believe that both parties, evaluation team and beneficiaries, should have similar key performance indicators.

The need to specify ways to use the SEC was also communicated. For example, one individual wrote:

Use and abuse of checklists
Increasingly checklists are becoming useful tools in development and that’s why they are so popular. Provided they are used appropriately, they can indeed be handy. Unfortunately, in program practice I have come across the frequent inappropriate use of checklists, in particular in cases where they are followed religiously, almost like a recipe book, without really understanding the complexity of what we are trying to achieve. Underlying seems to be an assumption that, as long as you tick the boxes of the checklist, you’ll be alright. Without further thought, checklists can become an excuse for not thinking any further.

Your checklist is comprehensive enough to look at many different factors that ultimately contribute to sustainability. The question though is, how will the user decide what to focus on? This will depend, to a large extent, on the intervention logic that determines the strategy for any development initiative. This guidance, in my opinion, is currently absent in the document.
Other themes that emerged within confusions included “food for thought” and “tensions,” both of which frequently occurred across items and will be further discussed later in this chapter (see cross-item analysis).

Errors/problems. As indicated in Figure 29, responses within errors and problems commonly related to “open questions and concerns.”

![Figure 29. Errors and Problems: Frequency of Emergent Themes](image)

While some of these were addressed within other items, the following types of issues emerged here:

The feeling I get is that systemic issues and concerns, which are at the heart of sustainability evaluation, are made to fit the criteria of merit, worth, and value. That’s not a problem per se, as long as there is space given to a reflection of what happens to merit, worth, and value once systems issues are fully integrated. Is the author of the checklist able to put her own cultural
tradition of evaluation in perspective and situate herself in relation to other traditions? That might need arguing out in the introduction a bit.

*I think it would be a mistake to attempt formulating a checklist that is universally applicable to any sustainability evaluation. There should be a core of universally applicable checklist items. But some tailoring of checklist items to fit each specific project is inevitable. And the checklist, including the core items, will probably have to be updated every five years or so. One more question comes to mind: What kind of training will be required for a person to be a checklist user for sustainability evaluation?*

Many responses regarding errors and problems were concerned with reoccurring issues related to language, length of the document, selectivity, food for thought, and tensions between specificity and generality. Additionally, specific spelling errors and formatting problems were highlighted. On the more uplifting side, some individuals referred to their use of the checklist. For example, one person remarked,

*It was very helpful for planning my most recent M&E plan.*

Finally, a few respondents raised issues about who is intended or best suited to use the checklist (i.e., “users”). Two exemplary responses highlighting the difference in the perceptions of respondents are displayed below:

*It's a list for evaluators. Use of evaluation "jargon" may put off others.*

*The list appears to be more relevant to program directors than evaluators.*

*Or at least, equally relevant to both. I would suggest the program directors*
and evaluators should discuss the checklist together as the ideas the list raises are pertinent to the sustainability of a program itself as much as the evaluation of its sustainability.

**Strengths.** As shown in Figure 30, respondents also referred to the SEC’s strengths. Comprehensiveness, structure, content, contribution, concepts, and clarity were the most frequently emerging themes. Less frequently occurring themes in terms of strengths were food for thought, generality, the holistic nature, use, detail, and tensions.

![Pie chart showing the frequency of emergent themes](image)

**Figure 30.** Strengths: Frequency of Emergent Themes

“Comprehensiveness” was expressed in statements like this:

*It's comprehensive. From my point of view, this is the main strength that a checklist should have, and this checklist has it.*
It is broad-based, indicating that sustainability itself is complex; its assessment follows along in complexity. I am very pleased you are doing this work.

It is comprehensive, includes the multiple dimensions of sustainability.

It touches all areas that border on sustainability.

Also, the “structure” was perceived as good by several respondents:

The systematic inquiry into the major components and the cross referencing within the checklist.

It is very logical, logical flow is present, technically is very sound, easy to follow and almost all checklists can be used as draft ToR [terms of reference, similar to request for proposals] for project/program sustainability evaluation

Furthermore, a number of respondents identified content as a specific strength, including specific sections and subsections, the glossary, and reference list. Below is an example response classified under this theme:

The question regarding "determine if and how the evaluation will be evaluated (metaevaluation)" is valuable, as most of the time, we do evaluation but we do not evaluate how our evaluation has been conducted. This is the best part, in my view.

Other individuals pointed out that the checklist is a “contribution” to the field and for practitioners. For example:
This will be a really excellent contribution to the field, and I look forward to seeing the final version.

It is the first evaluation checklist focused on sustainability as far as I know. This is a very important endeavor.

The differentiation between evaluation for sustainability and evaluation of sustainability is a new clarifier to evaluation.

Moreover, “concepts” used within the checklist, such as the differentiation between evaluation of/for sustainability and definitional properties of evaluation were perceived as strengths by some individuals:

The fact that (unlike a lot of the US literature), it isn't focused on the notion that sustainability is about routinization of processes and procedures, but it is linked in with a dynamic, turbulent, and often unknowable environment and context.

The separation of merit, value, and importance...

The strengths of the checklist are that it provides a good contextual framework for focusing on evaluation of sustainability as opposed to evaluation of short-term outcomes, process evaluations, etc.

To distinguish between evaluation of sustainability and evaluation for sustainability I find most useful.
A few individuals felt that the SEC provided maximum "clarity" despite of the complex issues communicated within. For example, one respondent offered this observation:

*The SEC is very thorough and clear and breaks it all down. The way the checklist limits/guides the evaluator's focus is strength.*

Still, and as indicated within earlier discussed items, the SEC provides "food for thought," a characteristic that was perceived as a strength. For example:

*It stimulates perspectives on the different dimensions of sustainability.*

*I see the strength primarily in that it invites the evaluators to place the evaluation in a broader context and suggests points to consider which they may not originally have thought of. Time and resource limitations may cause not all aspects to be included in the final scope of the evaluation, but it will be clearer to the team what has been omitted.*

*It is really thorough and a good tool.*

Also the "generality" and "holistic nature" of the SEC were thought of as positive attributes by some individuals:

*It can be used in different contexts, and serves as a good guide for sustainability evaluation.*

*The attempt to maintain a holistic understanding of sustainability incorporating human, social (including issues of justice, equity and diversity), economic, and environmental (to some extent) dimensions is laudable and potentially helpful if the key points about politics, exclusion, and climate change can be better integrated.*
The strengths of this checklist is its holistica, as it has covered almost every dimension which can come into one's mind.

A few respondents suggested that they would use the checklist or that it would be useful for specific purposes:

I would use the checklist part A and part B if I were designing an evaluation, especially if I were asked to be engaged at project inception (where evaluation should start, but usually doesn't!). It's an excellent guide.

Having a checklist, which may or may not be used to assess sustainability, can help in the initial project/program design in clearly identifying issues and activities that need to be included to begin laying the foundation for sustainability.

Furthermore, the "detail" provided in the checklist as well as "overcoming the tensions between brevity and the level of conciseness" were complimented.

Suggestions. The final item included suggestions for improving the SEC. As indicated in Figure 31, the most frequently occurring themes were requests for "clarifications" and "specifications" within the SEC (cf., omissions), followed by requests to improve the utility.

Recommendations related to "utility" are evident in suggestions to "simplify" or "shorten." Some of the responses related specifically to improving the formatting and structure, adding supplementary materials (e.g., examples, cases, tables, etc.), and
developing of a user's guide. The theme "user guide" was used to code responses like this one:

*The checklist is very rich and condensed. In order to promote its use in evaluation and encourage sustainability evaluations, it would be useful to have an introductory section that emphasizes the purpose and main strengths of the list as well as the structure.*

*You could possibly provide a short description of your methodology and format at the front.*

![Figure 31. Suggestions: Frequency of Emergent Themes](image)

Again, language was frequently mentioned and strongly related to the "utility" category. The question is whether to remove evaluation-specific terminology or to imbue this subject-specific terminology throughout the SEC by simplifying the language in general and including the terms in brackets. Additionally requests for
new subdimensions were reaffirmed here, as well as the need to learn from checklist use.

**Cross-item Analysis**

Results from the cross-case analysis were used to analyze responses across items. Figure 32 displays a recategorization of emergent themes from the cross-case analysis. Themes illustrated in circles represent clusters of specific relevance to a given item. For example, specifications, clarifications, new dimensions, etc., are issues that are logically related to omissions. These are also issues that were frequently mentioned within omissions, but have occurred within other items (e.g., within errors and problems) as well. Similarly, overlaps, unnecessary parts/sections, reconsiderations, selectivity, and clarity are of particular relevance to commissions. The same logic applies to the other clusters presented above the items.

Arrows represent linkages between clusters, that is, they signify potential overlaps between clusters. For example, the clusters associated with omissions, commissions, confusions, and errors/problems are all interconnected as they present weaknesses (dotted lines). The strength cluster is interrelated with each cluster that presents weaknesses to indicate potential inconsistency between strengths and weaknesses (solid lines). Finally, the suggestions cluster overlaps the strengths cluster, because these recommendations have to be considered in terms of core strengths of the SEC and their relationship to weaknesses.
Tensions between generality and specificity, necessity and sufficiency, and complexity and simplicity run across items. Within omissions, respondents suggested that the checklist might be “too” complete and that the SEC included an overwhelming amount of detail. Within commissions, tensions were exemplified by the argument that stability and the ability to adjust to changing contexts are contradictory. Within errors and problems, tensions occurred in terms of having a universal checklist that has to be adapted to each specific context.

Figure 32. Cross-item Analysis
However, these tensions are also inherent in contradictory perceptions about the checklist. Some individuals thought that the SEC provided too little detail, while others thought that it had too much detail. Some people thought that the SEC is too simple while others found it too complex. And yet others thought that the idea of bridging the simple notion of “sustainability” with the broader concept “sustainable development” would be conceptually faulty, while others expressed that this would be a useful and unique contribution.

These tensions also emerged within the interviews. One interviewee described these tensions as a dilemma that might be resolved by providing an overview in the beginning of the SEC with additional details provided later on in the document. The request for such a division was also evident in other responses and is reflected in the cross-case analysis.

These tensions are also linked to the notion of “food for thought.” As shown in Figure 32 above, this theme was recategorized as a strength, because the critical thinking spurred by reviewing the checklist, which was evident in many comments, is important to evaluation. The notion of the checklist as a catalyst for deeper thinking about sustainability was also expressed in the interviews. One interviewee, for example, linked the SEC to the notion of a heuristic which “is something that makes you think.” As sustainability might be established via innovation in some contexts, through routinization in others, and by way of adaptability in yet others, the SEC could serve as a thinking tool.
Another interviewee who applied the checklist to one of his international development projects said, “What more can you expect than to stimulate thought processes? You are not going to get the same product every time, because that person has a lot more baggage in their head before they even pick this thing up.”

In the following, findings are discussed across sections, that is, by key strengths, weaknesses, and recommendations.

*Cross-section Analysis*

Contrasting the results within and between items, it is evident that both strengths and weaknesses exist, but that several issues are regarded as strengths by some individuals and as weaknesses by others. However, a relationship between these opposing perceptions and respondents’ levels of experience or education could not be established.

The comprehensiveness of the SEC was perceived as the predominant strength, while most pressing shortcomings related to utility, as evident in comments about the need for specifications, clarifications, reconsiderations, confusions, and other open concerns.

As a result, the need for new dimensions or subitems expressed under omissions must be considered in terms of suggested commissions. Both omissions and commissions need to be considered in terms of language, concepts, detail, and structure to minimize confusions. Other emergent problems have to be rethought in terms of the three aforementioned items.
Of special interest are convergences between the weakness clusters and the strengths cluster. As evident from the previous discussion, several responses contradict one another. While some individuals find the comprehensive nature of the SEC and the corresponding level of generality and detail as strengths of the checklists, several respondents suggested the opposite. Decisions about adding/removing checklist content, as well as about restructuring will have to be based on item-by-item consideration during checklist improvement.

Triangulation

Comparing the findings from the qualitative responses in relation to the rating items, it becomes evident that most weaknesses reflected utility, not accuracy, concerns. In essence, the lowest-ranking items on the utility scale were the most often discussed weaknesses. These included

- feasibility
- concreteness
- conciseness
- ease of following the SEC
- ease of applying the SEC.

Concerns that did relate to accuracy usually related to methodological concerns that would not be sufficiently addressed in the SEC. These were not assessed within the rating scale, but communicated via items that respondents could
add to the questionnaire as well as in the interviews. That is, a few respondents indicated that more information is needed about data sources, collection, and analysis—as well as related reliability concerns. The lack of methodological guidance was also referred to in one of the interviews, because these concerns would be unique in sustainability evaluation which commonly relies on prospective analyses “where people make heroic assumptions” instead of building on earlier projects and findings. As such, sustainability analysis would often just be an arithmetic exercise that is based on tenuous grounds. This could be strengthened in the SEC.

Accuracy items that were assessed in the scale did not commonly emerge as issues in the open-ended responses. The only item that was mentioned was the “identification of information sources.” One interviewee pointed out that the checkpoint mirroring this item was not specific enough (see interview summary #10).

The strengths of the SEC that were reflected in the qualitative responses largely converge with both the highly rated utility and accuracy items. For utility, there is general agreement that the SEC is important and relevant to international development. Despite the problems with the conciseness and concreteness, the SEC has been found “very, very useful.” For example, one questionnaire respondent stated,

I used the checklist to help plan a new M&E system for a new project, and I found it a most helpful guide and checklist. I developed a very complete plan that I think will be useful for internal use, but had to simplify it and condense it significantly for use in the proposal.
The SEC would also aid evaluation in different cultural contexts. One respondent said,

*My role in evaluation is mostly in program redirection and policy. It helps to do a cross-cultural (unicultural evaluation) where the tools of evaluation transcend cultural diversity. Yours is such a good tool.*

Also in relation to the accuracy scales, strengths expressed in open-ended responses reflected agreement that sustainability evaluations could be planned with the checklist and that the checklist is complete, considers important aspects of sustainability evaluation, supports the identification of relevant criteria, and facilitates metaevaluation.

One issue that appeared as a somewhat divergent theme emerged from interviews (see interview summaries #6 and 9) and e-mail responses, and relates to the systemic nature of sustainability. In essence, these respondents communicated that the sustainability definitions in the SEC are limiting or nonsystemic. One e-mail respondent said,

*I think the definition of sustainability is too narrow. I prefer the perspective of ecosystem and complexity scholars like C. S. Holling (Panarchy) that sustainability is defined by the ability to deal with turbulence. The definition you are using is quite static. As a result, the checklist is based on a traditional, largely static view of sustainability. It lacks a complexity understanding or perspective. A more adaptive perspective would broaden criteria. That may be more than you can do at this stage, but you should acknowledge it as a major limitation.*
An interviewee suggested that in bringing the systems perspective into the checklist would require consideration of interrelations, perspectives, and boundaries.

Summary and Implications for Checklist Improvement

Nonetheless, the SEC was characterized as valid and useful for its intended purpose and based on the findings presented above, several adjustments to the checklist are necessary. Aspects to be addressed in the short-term include distinct specifications and clarifications; language simplification; and the addition of a more descriptive introduction that addresses key concepts used within the checklist, elaborates intended uses and users, and exemplifies how the SEC should be used. Of special significance is the need for a summary table that can be used as a “quick guide” and refer users to more detailed discussions where necessary.

In the longer term, the checklist should be continuously improved and supplemented with clear examples, cases, and—potentially—training materials. Also, the systemic nature of the checklist content has to be better addressed. The checklist needs to be made interactive by developing a Web-based format in which parts, sections, subsections, and dimensions are interlinked and connectivity to key terminology and resources is provided.

The next and final chapter provides more detail to changes made to the checklist at this time and proposes a research agenda for addressing other salient
issues in the future. The final chapter will elaborate on the limitations as well as the contributions of this dissertation.
VI. DISCUSSION AND CONCLUSIONS

This dissertation set forth to develop and validate a checklist (the SEC) to facilitate evaluations of sustainability for sustainability, thus bridging the gap from unidirectional evaluation to bidirectional and potentially multidimensional sustainability evaluation. To build the framework and compile the information necessary to alleviate (not solve) problems in sustainability evaluation, a literature review was presented in Chapters I and II. Based on the pertinent literature in evaluation, sustainability, and sustainable development, a checklist was developed (see Chapter III) using guidelines presented by Stufflebeam (2000), Scriven (2007), and Hales and colleagues (2007). This checklist was then iterated through multiple feedback loops (i.e., AEA, 2007 panel and roundtable; pretest and pilot test with IDPE students; Evaluation Café Think Aloud). To garner critical, detailed feedback on the SEC and its validity and utility from experts and potential users in varying development contexts, a nonexperimental, exploratory, primarily qualitative study was conducted. The previous chapter (Chapter V) presented the results of the SEC study and suggested a need for revising the checklist in the short and long-term.

This concluding chapter presents revisions made to the SEC as a result of the extensive feedback obtained in the study, discusses limitations and their implications for future research, describes the contributions of the dissertation to development
practitioners and the field of evaluation, and concludes with future directions. Figure 33 provides a chapter overview.

**Figure 33. Chapter VI Overview**

**Revisions**

In light of the findings presented in Chapter V, the SEC was revised (see Appendix L). Modifications include both changes to the content and format/structure (see Table 14). This section also discusses what has not been addressed to date.

Specifications contain, for example, the addition of evaluation purposes based on Patton's six-dimensional model (2008, forthcoming), the incorporation of politics
and power concerns in evaluation and for the evaluand (ibid.), the integration of ethical concerns (Morris, 2008; Stufflebeam, 2007), and consideration of competencies that may be required in an evaluation team among others.

Table 14

Changes to Content and Format/Structure

<table>
<thead>
<tr>
<th>Main findings</th>
<th>Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of specifications</td>
<td>• Elaboration of purposes</td>
</tr>
<tr>
<td></td>
<td>• Integration of politics, power, and ethics</td>
</tr>
<tr>
<td></td>
<td>• Illumination of evaluator competencies</td>
</tr>
<tr>
<td>Need for clarifications</td>
<td>• Expansion of SEC introduction</td>
</tr>
<tr>
<td></td>
<td>• Amplification of timeframes, metaevaluation and evaluation standards, sustainability dimensions, stakeholders and impactees, and procedural guidance</td>
</tr>
<tr>
<td>Need for “quick guides”</td>
<td>• Addition of a summary table and a flow chart</td>
</tr>
<tr>
<td></td>
<td>• Provision of figures and tables</td>
</tr>
<tr>
<td>Structure/Format</td>
<td>• Reordering of criteria</td>
</tr>
<tr>
<td></td>
<td>• Reclassification of risks from worth to significance</td>
</tr>
<tr>
<td>Simplifications</td>
<td>• Adjustments to language throughout the SEC</td>
</tr>
</tbody>
</table>

Clarifications were made throughout the checklist. In Section 1, the timeframe for evaluation was separated by respective evaluation approaches used (i.e., prospective versus retrospective evaluation), metaevaluation was linked to the Program Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 1994), and reference was made to the availability of evaluation standards in many international and national organizations and associations. In Section 2, sustainability dimensions were distinguished by evaluation of sustainability and evaluation for sustainability. The stakeholder and impactee checkpoint was separated
to uniquely address stakeholders and impactees by internal and external groups. In Section 3, the general procedures of evaluation were rewritten to include more procedural guidance. For example, a modified version of Davidson’s (2005) strategies to determine weights of importance has been included.

As there appeared to be much disagreement as to whether checkpoints, sections, or whole parts should be excluded, and the comprehensiveness of the checklist was named as the primary strength, not many items were removed from the SEC. At the same time, efforts were made to keep the length of the SEC to a minimum in order to thwart overburdening or checklist anxiety. In contrast to evaluation anxiety, defined as fear “provoked by the prospect, imagined possibility, or occurrence of an evaluation” (Scriven, 1991, p. 145), checklist anxiety was expressed by respondents in terms of the complexity of the topic area and tasks involved in conducting evaluation, the length of the checklist and the academic prose.

To thwart this anxiety, the introduction was expanded to clarify the objectives of the SEC, intended users and associated uses, characteristics, concepts, and to provide more guidance in general. Specifically, the “heuristic” nature of the checklist was stressed in the characteristics section (Ulrich, 1991). It was also clarified that the SEC is not intended to function as a tool that must be adhered to in all cases and contexts. Instead the user is urged to use common sense in applying the tool. While each point within the checklist should be considered, not all are applicable in all cases. To minimize confusion about the required level of detail and adherence and to increase the checklists feasibility, a summary table (see Table 15) has been provided
where users can indicate whether a checkpoint is applicable or not applicable in a given situation (Hales et al, 2007).

Table 15
SEC Summary Table: Quick Guide

<table>
<thead>
<tr>
<th>Part A – General considerations in evaluation</th>
<th>Clear</th>
<th>Needs consideration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1: Grounding the EVALUATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direction of the evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User of findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose(s) of the evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roles of the evaluation team</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Composition of the evaluation team</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timeframe under evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Key questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metaevaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissemination of findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 2: About the EVALUAND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The evaluation object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Components of the evaluation object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustainability dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local historical Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stakeholders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impactees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 3: General procedures for evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collect data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synthesize</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 15—continued

<table>
<thead>
<tr>
<th>Part B – Criteria</th>
<th>Not relevant</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 6: Significance: Is the continuation of the evaluand important?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs for human sustainability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs for social sustainability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs for economic sustainability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs for environmental sustainability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scope and duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risks and vulnerability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 4: Merit (Quality): What are the properties which define good sustainability without consideration of cost?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of evidence from monitoring and evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appreciation of knowledge, skills, abilities, competencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership competencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizational characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaboration/involvement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding the community and its environmental context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal orientation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive and negative impacts over time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 5: Worth: Is the continuation of the evaluand or its outcomes worth the costs that accrue now and potentially will accrue in the future?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time at which costs/resources are accrued</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stakeholders and impactees, to whom monetary and nonmonetary costs accrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facets of cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific costs or resource use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource renewal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Furthermore, the introduction of the SEC includes a simple flowchart (see Figure 34) intended to facilitate checklist use and maximize checklist feasibility by aiding decision making about what sections of the SEC should be consulted.
Evaluation with an emphasis on sustainability requested

Don't use this checklist

Yes

Are you clear about the general considerations in evaluation?

No

Not sure

Check Section 1

Yes

Do you understand the object under evaluation in terms of sustainability?

No

Not sure

Check Section 2

Yes

Are you clear about the general logic of evaluation?

No

Not sure

Check Section 3

Yes

Are you clear what to consider to learn whether the evaluation object is important?

No

Not sure

Check Section 4

Yes

Are you clear about the properties that define good sustainability?

No

Not sure

Check Section 5

Yes

Are you clear about what to consider to learn whether continuation is worth the costs that accrue now and in the future?

No

Not sure

Check Section 6

Figure 34. SEC Flow Chart
Finally, the SEC has been imbued with a few figures and tables intended to facilitate critical thinking and checklist use. To better illustrate the notion of a life cycle (e.g., Vigon, Tolle, Cornaby, Latham, Harrison, Boguski, Hunt, Sellers, & USEPA Risk Reduction Engineering Laboratory, 1994), prospective evaluation for sustainability and retrospective evaluation of sustainability has been depicted within a program life cycle model. To consider the past, current, and projected context of the evaluand, a table was included for the local historical context. Davidson’s (2005) detriment-benefits matrix was adapted and incorporated in Section 4, Criteria of Significance to facilitate an understanding of assessing needs and risks, as well as an example to illuminate the weighting process in evaluation.

Changes to Format and Structure

Formatting and structure was clarified. First, the section on Criteria of Significance was moved before the Criteria of Merit section to better comply with the logical order of evaluation. Risks, originally captured within the Criteria of Worth were moved under the criteria of significance, also to better support the logic presented in the checklist.

Few individuals suggested incorporating Part B into Part A, Section 3 and to remove the distinction between process-oriented and outcome-oriented criteria of merit. While this change could be of advantage to some, it was not implemented here, because this would confine the most important aspect of the SEC and furthermore unnecessarily complicate this already difficult section.
Omissions: That which has not been done

The systemic nature of the checklist content has to be better addressed, that is, the checklist needs to be made interactive by developing a Web-based format, where parts, sections, sub-sections, and dimensions are interlinked and connectivity to key terminology and resources is provided. Initial efforts in this direction have been taken. A Web domain is available and the final product of this dissertation will be posted with cross-linkages and interrelationships illuminating the systemic nature of the sustainability evaluation.

Furthermore, several individuals requested the inclusion of specific examples and cases. This was not possible at this time because it requires application. The intention is that users will volunteer examples and cases to expand on the knowledge base contained in the checklist.

Another issue is the increased utility for people with disabilities. While the font size has been increased to 12 points throughout the document, additional adaptations may be of importance to assure user-friendliness for people with disabilities. To do that requires follow-up with individuals from the group.

Moreover, the structure and sequencing has to be more illuminated. While initial adjustments were made, application will show what works for whom and how the checklist can be modified to best serve potential users. This problem also parallels the language constraints. SEC use will likely lead to additional changes and refinements.
Finally, potential overlaps may exist due to the nature of the checklist. As users are not forced to apply each individual step, some issues could appear redundant to those who apply the complete SEC. Checklist use will show whether this is problematic or can be modified to improve the checklist in the long run.

Limitations and Outlook

As with most dissertations, the work presented here has several limitations, including shortcomings in capturing all available literature, drawbacks inherent in the checklist development, and the boundaries of the validation method. Many of these, however, provide guidance for a prospective research agenda in sustainability evaluation.

The literature included in this dissertation captured a wide variety of different perspectives and approaches to look at sustainability. Yet, it only appears to have scratched the tip of the iceberg of available resources and current thinking about sustainability, sustainable development, and sustainability evaluation. The fact that conceptions about sustainability are often ambiguous and varied exacerbates this drawback. Within a dissertation, even if conducted over a long period of time, it is hardly possible to go beyond the most prominent works. Future work on the topic can dig deeper in terms of each unique checkpoint established in the SEC. Of special relevance is an in-depth exploration of the systems literature to capture the complexities of sustainability and move beyond the relative static representation in
the SEC and dominant literature on sustainability. Concepts to be explored include theories of complex adaptive systems as presented by Marion (1999) or Gunderson and Holling (2002), to name but a few. Additionally, the varying perceptions of sustainability in Europe could be better explored, yet require time and resources to access all relevant information as well as overcome potential language constraints.

The checklist’s development was thwarted by the fact that it has been done exclusively in an ivory tower setting. That is, development was not supported by continuous engagement in the international development programs or organizations. While my participation in the EASY ECO and the United Nation’s discussion in consultation for the 2008 AMR were insightful and led to further readings about the sustainability concerns we face, practical experience will illuminate feasibility and practicality of the tool. That is, although many examples and means are available in the literature, a collaborative effort within an organization may have led to a more practical tool in the first place. One of the core criticisms in the study was that the checklist is too technical and academic. The extensive feedback provided by experts and practitioners relieved this problem to some extent. Yet, the checklist will have to be tested and further revised according to practitioner feedback and future studies.

Checklist validation was limited to the assessment of content validity and perceived utility. Future efforts must emphasize checklist application and respective implications for reviewing and refining the work. While the checklist has been applied by several respondents, the feedback obtained was not tailored to this specific task, because the study set forth to investigate the content and usefulness of the
checklist specifically. On the upside, those who did apply the checklist confirmed its utility to practitioners in the international development arena. Further, the study was not sponsored, thus respondents who did apply the checklist did not benefit any more from this research than the ones who did not apply the checklist.

At this time, the test version was disseminated and the revised version of the SEC will be circulated in the near future. While all volunteers received the test version, the revised checklist will only be provided to those individuals who provided their feedback on the checklist. Feedback from these potential users might guide further redevelopment of the SEC. Additionally, efforts are in place to integrate the SEC into a Web site on sustainability evaluation. This resource will be freely available for anyone interested and include links to further readings, methodologies, tools, and resources provided on the Internet. If funding for future development should become available, the checklist content will be transformed into training modules that enhance checklist use, thus evaluation implementation. Moreover, the SEC will be submitted for review to the EC Checklist Project for potential dissemination via the EC Web site.

While the research on checklist use is limited to date, several issues deserve future study, including the following questions:

1. When and for what tasks is evaluation checklist use appropriate? Hales and colleagues (2007) suggest that checklist overuse may become a hindrance in clinical settings. This can also be the case in evaluation efforts.
2. What level of complexity is appropriate for intended users? How can different levels be established, disseminated, and promoted without comprising the checklist task at hand? Because checklists are intended to fulfill a support and error-reduction function for their users, it is important to minimize unnecessary levels of complexity, and thus increase levels of reliability in checklist use (cf., Hales et al, 2007).

Finally, Scriven (2007) argues that checklists are short of being theories or methodologies although they incorporate complex sets of assumptions. Future research on the SEC may contribute to the development of a theory of sustainability evaluation.

Contributions to Evaluation Theory and Practice

Despite its limitations, this dissertation contributes to evaluation and the (international) development field. First and foremost, this dissertation led to the development of a new tool, the SEC. This tool was presented at numerous conferences and was published in its initial form as a book chapter (Schröter, 2008a). Moreover, the draft SEC was disseminated to all individuals who indicated an interest in the study and was immediately used by some individuals, indicating potential of the revised version.

Second, the SEC was evaluated by a heterogeneous group of experts and practitioners in the (international) development arena. Being at the core of sustainability evaluation, these individuals offered extensive feedback, much of
which is integrated into the product of this dissertation. Participants will also receive a revised and improved version of the checklist to enable further discourse that leads continual refinements of the SEC. The dispersed dissemination of the SEC also enables use, though application is not warranted.

Third, this dissertation details a methodology for validating checklists in their initial stage. This methodology, though limited to some extent, may provide ideas for others who are developing new tools for evaluation and intend to validate that work.

Concluding Remarks and Future Directions

Considering increasing attention regarding sustainable development worldwide, concerns about program sustainability, and continuation of positive outcomes and impacts for improving human, social, economic development within the constraints of the natural environment, this dissertation set forth to develop a tool that bridges the gaps between evaluations of sustainability and evaluations for sustainability. While the complexities of the issues addressed within this dissertation could only be scratched on the surface, extensive feedback on the tool allowed for improvements and supports the idea that the checklist can and will be used widely. Application and extended research, specifically as it relates to the systemic nature of the topic, will further enhance the SEC and its utility for both evaluators and development practitioners to assure that sustained programming positively affects sustainable development.
REFERENCES


APPENDICES

Appendix A. Acronyms

This appendix enumerates the many acronyms that are used throughout this dissertation. Latin abbreviations (such as cf., e.g., i.e., viz., vs., and etc.) as well as commonly used abbreviations (e.g., ed./eds., U.S., or USD) are not itemized here, as they are assumed to be common knowledge. If abbreviated terms comprise foreign language text, the word or phrases are being translated in parenthesis to increase the understanding.

- AEA: American Evaluation Association
- AMR: Annual Ministerial Review
- APA: American Psychological Association
- ATE: Advanced Technological Education
- BMBF: Bundesministerium für Bildung und Forschung (German: Federal Ministry of Education and Research)
- BRD: beyond reasonable doubt
- CGIAR: Consulting Group on International Agricultural Research
- CIPP: Context Input Process Product
- CFO: Code of federal regulations
- CSD: Commission on Sustainable Development
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTA</td>
<td>Technical Centre for Agricultural and Rural Cooperation ACP-EU</td>
</tr>
<tr>
<td>DAC</td>
<td>Development Assistance Committee</td>
</tr>
<tr>
<td>DANIDA</td>
<td>Danish International Development Agency</td>
</tr>
<tr>
<td>DeGEval</td>
<td>Deutsche Gesellschaft für Evaluation (German: German Society for Evaluation)</td>
</tr>
<tr>
<td>DFG</td>
<td>Deutsche Forschungsgesellschaft (German: German Research Foundation)</td>
</tr>
<tr>
<td>DFID</td>
<td>UK Department for International Development</td>
</tr>
<tr>
<td>DG JRC</td>
<td>European Commission's Joint Research Center</td>
</tr>
<tr>
<td>EASY ECO</td>
<td>Evaluation of Sustainability: European Conferences and Training Courses</td>
</tr>
<tr>
<td>EC</td>
<td>The Evaluation Center</td>
</tr>
<tr>
<td>E-CDC</td>
<td>Evaluation – Checklist Development Checklist</td>
</tr>
<tr>
<td>ECOSOC</td>
<td>United Nations Economic and Social Council</td>
</tr>
<tr>
<td>EES</td>
<td>European Evaluation Society</td>
</tr>
<tr>
<td>EPA</td>
<td>Environmental Protection Agency (US)</td>
</tr>
<tr>
<td>ERA</td>
<td>European Research Area; evaluation of research by academic field</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>FAO</td>
<td>UN Food and Agriculture Organisation</td>
</tr>
<tr>
<td>FP7</td>
<td>Seventh Framework Programme</td>
</tr>
</tbody>
</table>
GAO General Accounting Office
GDP gross domestic product
GTZ Gesellschaft für technische Zusammenarbeit (German International Development Agency)
HSIRB Human Subjects Institutional Review Board
IDPE Interdisciplinary Ph.D. in Evaluation
IDPM Institute for development policy and management
IIED International Institute for Environment and Development
IISD International Institute for Sustainable Development
ISO International Organization for Standardization
KEC Key Evaluation Checklist
MDG Millennium Development Goals
NGO non-governmental organization
NSF National Science Foundation
NWS numeric weight and sum
ODI Oversees Development Institute
OECD Organisation for Economic Co-Operation and Development
RFP Request for proposals
RIT The Rochester Institute of Technology
SD Sustainable development
SEC Sustainability Evaluation Checklist
TIG Topical Interest Group
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToR</td>
<td>Terms of reference</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>UNCED</td>
<td>United Nations Conference on Environment and Development</td>
</tr>
<tr>
<td>UNDESA</td>
<td>United Nations Department of Economic and Social Affairs</td>
</tr>
<tr>
<td>UNDP</td>
<td>United Nations Development Programme</td>
</tr>
<tr>
<td>UNEP</td>
<td>United Nations Environment Programme</td>
</tr>
<tr>
<td>UNESCO</td>
<td>United Nations Educational, Scientific, and Cultural Organization</td>
</tr>
<tr>
<td>UNS</td>
<td>Unified National System</td>
</tr>
<tr>
<td>USAID</td>
<td>United States Agency for International Development</td>
</tr>
<tr>
<td>WCED</td>
<td>World Commission on Environment and Development</td>
</tr>
<tr>
<td>WMU</td>
<td>Western Michigan University</td>
</tr>
</tbody>
</table>
Appendix B. Sustainability Evaluation Checklist (Test Version)

NOTE: The checklist has been adjusted to fit the format of this dissertation. The reference list and table of contents have been excluded to prevent unnecessary information within an appendix. Furthermore, the checklist was formatted to the standard dissertation layout.

The title page of the checklist included the following note for participants:

This document consists of a brief introduction, the Sustainability Evaluation checklist (SEC), a glossary of terms, and references to key readings. The checklist presented here is a brief version (about 10 pages) intended for gathering feedback, while minimizing the amount of time needed to work through the details.

When reading the checklist, please think about the following questions:

1. What is missing from the checklist?
2. What components or checkpoints are not necessary?
3. Are there any other errors or problems that need to be addressed?
4. What, if anything, do you like about the checklist?
5. Do you have any suggestions for how to improve the checklist?

Thank you very much for your time and effort!
Introduction

*Indicates terms (words) explained in the glossary*

Purpose

The Sustainability Evaluation Checklist (SEC) is intended for use in planning and designing project and program evaluations of sustainability for sustainability within development contexts. As a mnemonic device, the checklist aims to reduce errors of omission and increase evaluation usefulness.

Intended users

The SEC is intended as a guide for practitioners who conduct internal or external evaluations of projects or programs. Experienced evaluators may benefit from the comprehensive nature of the checklist, while new evaluators or those with limited exposure to evaluation will find value in the instructional elements of the tool.

Characteristics

The checklist is built on Scriven’s (1982) and Fournier’s (1995) logic of evaluation and consists of two major parts: (A) general considerations and (B) criteria of merit, worth, and significance. The general considerations, Part A, are subdivided into three sections: (1) grounding the evaluation, (2) about the evaluand, and (3) general procedures for evaluation. In part (B), criteria for evaluating sustainability are distinguished by (4) merit, (5) worth, and (6) significance.

Key concepts

Sustainability

The capacity to exist (e.g., projects, programs, mankind) or continue (e.g., human, social, economic, and/or environmental benefits). For programs, this usually means existence (temporal durability) beyond termination of initial support via mechanisms that have been used to develop the program. In terms of continued benefits, it means that programming does not negatively impact human survival on earth.

Evaluation

The determination of merit (quality), worth (value), and significance (importance)

Evaluation of sustainability

The determination of the merit, worth, and significance of efforts to continue a given evaluand (i.e., evaluation object) beyond the removal of initial program resources: What is the level of sustainability of your evaluand? How well is the evaluand sustained?
Evaluation FOR sustainability

The determination of the merit, worth, and significance in maintaining, replicating, and exporting a given evaluand’s positive (un)intended outcomes and impacts under specific consideration of global sustainability issues. How well does the evaluand contribute to sustainable development efforts (human, social, economic, and environmental dimensions)?

Checklist A tool that guides evaluation efforts, also known as a framework for conducting evaluation

Sustainability Evaluation Checklist
(Test Version, March 2008)

Part A – General considerations in evaluation

The following checkpoints should be discussed among evaluation team members, the evaluation client, and key evaluation stakeholders to clarify information needs, resources, methodological decisions, required levels of detail, and evaluation management. They comprise general considerations in evaluation that have been specified for sustainability concerns. In addition, you may also want to consider general evaluation checklists available at http://www.wmich.edu/evalctr/checklists/checklistmenu.htm

Section 1: Grounding the EVALUATION

These general considerations are of major importance for planning the evaluation.

Determine the purpose of the evaluation: Why are you conducting the evaluation? How are findings to be used?

- Improvement-oriented (formative*): Is it the intent of the evaluation to improve the sustainability of an evaluand and/or its outcomes and impacts?
- Accountability oriented (summative*): Is it the intent of the evaluation to hold program leaders and staff accountable for the level of sustainability achieved with the available resources?
- Knowledge generation (ascriptive*): Is it the intent of the evaluation to generate knowledge about how sustainability is manifest within an evaluand?

Determine the direction of evaluation: Who asked for the evaluation?

- Bottom-up*: the evaluation begins on the ground at the local project level (i.e., is driven by middle range or grassroots actors, focusing on a specific project)
Top-down*: the evaluation begins on the highest level (e.g., is mandated by the funder). Sustainability may only be one dimension to consider among others (e.g., by the DAC criteria)

Clarify the role of the evaluator/evaluation team:
- Internal evaluation team (e.g., staff members, funders, donors)
- External evaluation team (e.g., independent consultants or other entity)
- Mixed (e.g., the evaluator as a critical friend)

Clarify audiences* to determine adequate means for disseminating findings from this evaluation:
- Who must receive the evaluation results for what purpose or use?
- Who else should receive the evaluation results, why?
- Who are the intended users of the evaluation?
- How should findings be disseminated to facilitate learning?

Define the timeframe under evaluation: In what stage of the lifecycle is the evaluand?
- Conceptualization and/or development: Sustainability is being considered prospectively
- Growth and/or maturation: The level and breath of sustainability are considered
- Reduction of initial funding resources: The stability of the evaluand and its outcomes and impacts are considered in terms of reduced resources or altered funding streams
- Termination of initial funding resources: The stability of the evaluand as well as of the breadth and depth of outcomes and impacts are considered once initial resources have been terminated and funding streams adjusted to the new situation
- After initial funding has ended: The stability of the evaluand as well as of the breadth and depth of outcomes and impacts are considered months after initial resources have been terminated and funding streams adjusted to the new situation (see termination of initial funding resources, the same considerations in a longer time frame)

Determine the type of evaluation
- Holistic*: the breadth and depth of sustainability of the whole evaluand is considered without separating parts and/or dimensions
- Analytic*: the breadth and depth of sustainability of components*, dimensions*, and/or the underlying theory* are considered in separate subevaluations before synthesizing these into conclusions about the whole evaluand
Determine the key questions to be answered by the evaluation.

- Evaluation FOR sustainability: How well does the evaluand contribute to sustainability in the long term? How well does the evaluand contribute to sustainable development efforts?
- Evaluation OF sustainability: What is the level of sustainability of the evaluand at this time? How well is the evaluand sustained?

Determine if and how the evaluation will be evaluated (metaevaluation)

- Consider the evaluation’s trustworthiness (i.e., validity), usefulness (i.e., utility), integrity (i.e., propriety), cost-effectiveness (i.e., feasibility; uses of evaluation results in relation to resources invested for conducting the evaluation), ethicality

Section 2: About the EVALUAND

This information is important for increasing understanding about the evaluand and its context.

Define the evaluand: What are you evaluating?

- Policy
- Proposal
- Program
- Project
- Process
- Activity
- Product
- Outcome
- Impact

Determine key stakeholders and impactees of the evaluand: Who is impacted?

- Recipients, participants, users
- Family member, friends, business partners, colleagues, community members at large, attentive audiences, consumers
- Individuals involved in implementing the project/program (e.g., staff)
- Funders, governments, NGO’s, etc.
- Alternative impactees: those who could have been impacted or protected from impact

Determine the reach of the evaluand: How far do impacts potentially reach?

- Consider local, national, regional, international impacts on people
- Consider potential changes over time

---

Supplemental information:

78 Commonly referred to as “beneficiaries.” The terms “recipient(s), participant(s), and user(s) are preferred here to assure that benefits are not falsely attributed.
Describe the nature of the evaluand in terms of sustainable development: What exactly is to be sustained by whom?

- Components: inputs, activities, outputs, outcomes, and potential (un)intended impacts of the evaluand
- Dimensions of the evaluand: human, social, economic, and environmental conditions and capacities; project-oriented, purpose/use-oriented, systems-oriented, and behavior-oriented sustainability
- Contextual influences: human (culture, beliefs, perceptions), social, economic (nature of economy and investment patterns), environmental (local/regional challenges), political (commitment), administrative, and technological specifics
- Resource availability: what natural, physical, human and monetary resources are available regardless of actual use?

These procedures lay out the working logic of evaluation.

Identify criteria: On what components or dimensions must the evaluand do well to be considered of good quality, value, and importance?

- Consult the list of criteria of merit, worth, and significance in Part B of this checklist
- Identify additional values and criteria via:
  - Needs and risk assessment: If no recent and/or valid needs or risk assessment is available, conduct one to determine human, social, economic, and environmental needs and risks that the evaluand ought to address
  - Organizational values: Are there any predetermined values specified by the organization(s) responsible for the evaluand?
  - Local and national policies: Are there any values of relevance to the evaluand that have been specified in local or national policies? Consider for example National Sustainable Development Strategies (NSDS), Poverty Reduction Strategy Papers (PRSPs), and others, if available
  - Legal requirements: What federal, national, or state laws and regulations are relevant across program boundaries? How could these laws/regulations affect sustainability?
  - Values and policies of the international community: Consider for example the Millennium Development Goals (MDG), the United Nation’s Development Assistance Framework (UNDAF), and Agenda 21

- Consider weighting the relative importance of the values of the evaluand: These weights can be numeric (e.g., quantitative scale from 1-5, 1-10, etc.) or
non-numeric (e.g., qualitative categories such as critical, important, or desirable). Consider:

- Severity of need for sustainability: the more urgent the need, the more important it is to address a given problem
- Severity of risk if need is not addressed: the higher the risk the more urgent the need is to be addressed

**Set standards: How well should the evaluand perform?**

- Determine indicators* that provide evidence of performance on a criterion: you may want to consult indicator frameworks that have been developed by the sustainable development community
- Determine minimum acceptable standards (bars’): Given the resources invested in the evaluand, is there a minimum level of sustainability that has to be achieved on a given component, dimension, and/or overall? Consider the following:
  - Critical dimensions: a bar could be set if there is a point on the dimension at which the evaluand would be unjustified.
  - Holistic barring*: is there a minimum acceptable performance of the evaluand in terms of sustainability overall (i.e., across dimensions)?
  - Global barring*: is there a need for setting a minimum level of combined or aggregated performance across all dimensions or components?
- Decide whether to grade* or rank* the performance on the criteria
  - Grading (rating): Assigning the evaluand or its components or dimensions to an ordered set of categories, with the order corresponding to a metric of merit, worth, and/or significance
  - Ranking: Placing the evaluand or its components or dimensions in an order of merit worth, and/or significance on the basis of their relative performance on a measurement or observation.
- Develop rubrics* that explain how performance on the criteria will be rated: Determine what it means to perform inadequately, adequately, or exceptionally (you may want to choose a different type of scale)

**Collect data (measure criteria/ performance) and compare with the standards:**

**How well did the evaluand perform?**

- Develop indicators to measure the performance on the criteria (values) of relevance to the evaluand
- Consider indicators of sustainability and sustainable development, if available and applicable in your specific context
- Consider time and space dimensions of sustainability
- Collect data from all relevant sources and assure broad participation
- If unavailable, develop instruments and collect evidence (facts) via document and literature reviews, observations, tests, questionnaires, interviews, focus groups and/or other methods of data collection
Decide whether to use scoring*, grading, ranking, and/or apportioning*. 

Synthesis*: Integrate data with the standards on criteria into a judgment of merit, worth, and/or significance

- Integrate the data (ratings or grading on criteria) with the standards
- Depending on whether you are doing a holistic or analytic evaluation, determine how the relative or absolute merit, worth, and/or significance will be determined
  - Grading and ranking provide unidimensional conclusions about components or dimensions
  - Profiling* provides multidimensional conclusions, usually depicting grades and comparative performance on components or criteria
- Consider scaling*: Do the evaluative conclusions warrant increasing the scale of the policy, project, or program? Will increased inputs result in increased or better outputs?
- Ensure that evaluative conclusions (claims) are legitimate
- Identify strengths and weaknesses of the evaluative conclusions

Part B – Criteria of sustainability

The following sections enumerate criteria of merit, worth, and significance in sustainability evaluation. In contrast to indicators, values/criteria of sustainability are those properties of an evaluation object that are part of good sustainability in a given context.

Section 4: Criteria of merit (quality)

What are the properties which define good sustainability without consideration of cost?

Process-oriented criteria: These criteria are especially useful in determining if the evaluand has the capacity for addressing sustainability needs (i.e., prospective considerations), but also to determine which of the evaluand’s elements (e.g., activities) persist after initial resources for the evaluand have been removed (i.e., retrospective considerations) or supplemented with other resources.

Leadership

- Championing: Capacity to promote sustainability, while preventing or mitigating negative impacts.
- Strategic plans for sustainability have been formulated and are in place: Alignment of activities and goals with sustainability needs of consumers and impactees (see criteria of significance)
  - People-centered approach
  - Strong political commitment
  - Consensus and long-term vision
  - Shared strategic and pragmatic vision
- Effective participation
- Realistic and flexible targets
- Succession planning
  - Policy support measures: alignment of intervention with national and/or international policies and priorities
  - Stability of staff: There is little turnover
  - Commitment to sustainable development
  - Balance between bureaucratic efficiency and democratic involvement

Organizational characteristics
  - Relevant knowledge, skills, and abilities of staff members
  - Diversified funding streams
  - Diversified activities
  - Participation, involvement, and integration
  - Institutionalization efforts are in place
  - Continuous monitoring of progress toward sustainability
  - Legal basis of the organization

Infrastructure: Reach, condition, and match of infrastructure in relation to program/project goals.
  - Adequacy of technology (e.g., communications, mobility)
  - Accessibility of people to be reached (e.g., roads, vehicles)
  - Stable electricity, if needed
  - Adequate waste treatment
  - Adequate facilities for activities, program support, etc.

Collaboration
  - Inclusion of relevant stakeholders and impactees
  - Clear communication (transparency): sharing vision
  - Linkages to other organizations/partners
  - Communication patterns among participants on the local, national, and international level of the evaluand and respective knowledge transfer (systemic support mechanisms)
  - Collective responsibility and accountability

Understanding the community and its environmental context: In addition to understanding the community’s human, social, economic, and environmental needs, consider in particular:
  - Respect (sensitivity) for the community’s tolerance for change (i.e., cultural enablers and inhibitors to sustainability)
  - Acceptability across impactees: Intra-generational equity
  - Political stability
Recognition and preservation of diversity

- Community linkages (social capital: willingness to share knowledge and information, help in handling everyday matters, and reinforcing social networks, solidarity)

Responsiveness

- Flexibility in addressing emergent needs within the realm of the organization's mission and priorities
- Ability to adjust to changing contexts
- Ability to adjust for unanticipated negative impacts and side effects (e.g., environmental degradation)
- Substitution of resources in contrast to usage of nonrenewable resources: waste reduction
- Internalization of costs
- Continuous adaptation of intervention to maximize benefits and minimize harm
- Intergenerational equity: consider potential harms of an intervention to future generations; inclusion of children and youth specifically
- Integrated renewal mechanisms

Use of evidence from monitoring and evaluation

- Staff have integrated the value of continuous improvement into their daily work.
- Consider human, social, economic, and environmental factors as well as ethics
- Evidence suggesting that the evaluand should and/or can be continued
- Evidence suggesting that the evaluand can be improved to meet sustainability requirements
- There is evidence that continued factors (e.g., activities or outcomes) are satisfactory, that is, no less adequate than acceptable.
- There is evidence that factors that have shown to be detrimental or insufficiently have been discontinued or have been improved

Outcome-oriented criteria: These criteria might be especially useful in determining if the evaluand has the capacity for sustainability (i.e., prospective considerations) or which outcomes have been sustained to date (i.e., retrospective considerations). Consider all impacts, anticipated or not.

Goal orientation (key issues in sustainable development):

- Consideration of the whole system and its parts
- Consideration of human, social, economic, and environmental sub-systems (holistic science and appropriate technology)
Human subsystems: consider for example developing full human potential; cultural, moral and spiritual sensitivity; self-determination; population growth

Social subsystems: consider for example gender sensitivity; social justice

Economic subsystems: consider for example institutional viability; viable, sound and broad-based economic

Environmental subsystem: consider ecological soundness

Consider positive and negative impacts of the evaluand on:

- Human dimension: consider for example health, education, poverty reduction; availability and quality of food
- Social dimension: consider for example equity; ethics; cultural beliefs, language, values; indigenous rights; community cohesion, stability, character, services, and social institutions; politics
- Economic dimension: consider for example access to and control over resources; infrastructure, institutions, tourism
- Environmental dimension: consider for example aesthetics (landscape analysis); archaeology and heritage; quality of air and water; level of exposure to risk, hazard, noise, dust; the local ecological condition on which life depends
- Cross-dimensional impacts: consider for example intergenerational and intra-generational equity; empowerment via social mobilization, direct action, power or protest-based confrontation, economic and social production approaches, civic engagement, raising consciousness, building capacity by providing knowledge, skills, and positive experience/success; usefulness of outcomes to community

Section 5: Criteria of worth (value)

Is the continuation of the evaluand or its outcomes worth the costs that accrue now and in the future?

Available cost evaluation checklists provide guidelines for identifying and analyzing costs and benefits that can be useful in many evaluations. In evaluation of sustainability, however, the consideration of cost must not only reflect on benefits and costs but also entail current and future generations.

Time at which costs/resources are accrued

- During the general program life cycle: What monies and resources are, were, and could have been used during the program life cycle?
  - This information should be available from previous evaluations, monitoring activities, or program documentation based on which the sustainability evaluation has been deemed worth doing
  - Consider adequacy of financial, human, material, and other resources
After the termination of startup funding: what monies and resources are, were, and could have used to continue elements of the evaluand? These monies and resources show the capability of the evaluand to continue.

Time of future generations (e.g., 20 years later): what monies and resources are needed to continue important elements of the evaluand in the long run? These costs are those resources and capacities required for maintaining the evaluand and outcomes thereof over time.

Stakeholders and impactees, to whom monetary and nonmonetary costs accrue (check your list of consumers and impactees developed earlier to make sure that you do not forget anyone or any group of people or organization that is of importance here)

- Costs that accrue for those people involved in the programming. These costs includes payments/salaries, time, resources, but also personal costs (e.g., stress, time away from the family, etc.)
- Costs that accrue for those people that are impacted by the programming. These include intra-generational and intergenerational impactees and upstream stakeholders as well as alternative impactees.

Facets of cost (generally, monetary and nonmonetary are distinguished here):

- Actual costs to humans (monetary and nonmonetary costs that accrue to individuals)
- Actual costs to society (monetary and nonmonetary costs that accrue to groups of people, organizations, communities, etc)
- Global costs (e.g., economic and environmental costs)
- Opportunity costs*: the cost of not considering alternatives at each of the previously stated levels

Specific costs or resource use to consider include:

- Human resources
- Renewable and nonrenewable resources
- Tools and technologies used
- Infrastructure
- Recycling, waste management, and conservation
- The benefits from the evaluand are equal to or larger than the costs accrued

Risks*: Unawareness of risks may thwart sustainability. It is critical to be aware of potential risks to the sustainability OF the evaluand that threaten potential FOR long-term sustainability. What strategies are in place to thwart potential risks to successful continuation of relevant and successful components and dimensions of the evaluand?

- Flexibility to changes in the environment
- Cultural compatibility of activities
- Risk to human sustainability: e.g., overpopulation
- Risks to social sustainability: e.g., violence and social breakdown
- Risk to economic sustainability: e.g., crisis and shocks, balance of payments
- Risk to environmental sustainability: e.g., natural disasters (earthquakes, tsunamis, etc.), over-consumption, waste, etc.
- Risk to participants and program staff if the evaluand is or is not sustained

Section 6: Criteria of significance (importance)

Is the continuation of the evaluand important?

Needs of impactees (e.g., human, social, economic, and environmental needs)

- Human sustainability:
  - Basic material human needs: food, shelter, clothes, medical care, schooling/education, means of transportation and communication
  - Basic non-material human needs: health, order, safety, belongingness, creativity, identity, autonomy, togetherness, participation, self-fulfillment; realization of an individual's potential
- Need for social sustainability:
  - Social norms; community cohesion for mutual benefit; connectedness between groups of people; solidarity; tolerance, respect, compassion, patience, and honesty; discipline; commonly shared rules, laws, and information; equity; human rights across cultures, natures, gender, religions; equity; peace; participation in decision-making about planned interventions that affect people's lives; justice; accountability
  - Self-reliance: specifically mobilization of communities, local ownership in decision making, commitment of local resources
- Need for economic sustainability:
  - Economic benefits to recipients, midstream consumers, or downstream consumers and other impactees; reduced need for external assistance
- Need for environmental sustainability:
  - Environmental soundness of the intervention, its intended and unintended outcomes and impacts
  - Ecological balance
Spectrum:

- **Scope** (continuation of activities, service provision, or outputs AND replication, transfer, or export of the evaluand): consider numbers and types of activities, services, and outputs as well as number of recipients/participants

- **Scale** (level of sustainability): consider routinization and institutionalization of programs
## Glossary of Checklist Terms

<table>
<thead>
<tr>
<th>A</th>
<th>Activity/activities</th>
<th>Actions that are assumed by an evaluation object to achieve goals (see Frechtling, 2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Activity/activities</td>
<td>Actions that are assumed by an evaluation object to achieve goals (see Frechtling, 2007)</td>
</tr>
<tr>
<td></td>
<td>Analytic evaluation</td>
<td>Components, dimensions, and/or the underlying theory of the evaluand are considered separately prior to synthesizing these 'subevaluations' into conclusions about the whole evaluand</td>
</tr>
<tr>
<td></td>
<td>Apportioning</td>
<td>Allocation or distribution: Dividing a given, often finite, quantity of valued resources between competing demands.</td>
</tr>
<tr>
<td></td>
<td>Ascriptive evaluation</td>
<td>Evaluation for the purpose of knowledge gain (for the sake of it)</td>
</tr>
<tr>
<td></td>
<td>Audience</td>
<td>Users of the evaluation; those who should receive reports, presentations, workshops, debriefings, etc</td>
</tr>
<tr>
<td>B</td>
<td>Bar</td>
<td>Minimum acceptable standard; an evalutive operation where minimum levels of performance are set, or required, on specific dimensions or components, performance below which cannot be compensated for by better performance on other dimensions. Failure to 'clear' a bar means 'failure' of the evaluand</td>
</tr>
<tr>
<td></td>
<td>Behavior-oriented sustainability</td>
<td>The target group or project holder has problem-solving capacities to adequately and flexibly adapt to changing environmental conditions (after Stockmann in Caspari, 2004, pp. 67-68)</td>
</tr>
<tr>
<td></td>
<td>Benefit</td>
<td>Positive outcome or impact</td>
</tr>
<tr>
<td></td>
<td>Bottom-up evaluation</td>
<td>The evaluation is initiated on the grassroots level</td>
</tr>
<tr>
<td>C</td>
<td>Checklist</td>
<td>A tool to plan and design evaluation</td>
</tr>
<tr>
<td></td>
<td>Component evaluation</td>
<td>Each part of the evaluation object (e.g., inputs, activities, outputs, outcomes) is evaluated separately. The resulting subevaluations are then integrated into overall conclusions about the evaluand (see Davidson, 2005a).</td>
</tr>
<tr>
<td></td>
<td>Criteria</td>
<td>Properties that are part of the concept of a “good X;” they are definitionally connected with the evaluand (see Scriven, 2007)</td>
</tr>
<tr>
<td>D</td>
<td>Dimensional evaluation</td>
<td>Merit, worth, and significance are considered for facets that permeate the whole evaluand; in sustainability evaluation, these facets include human, social, economic, and environmental dimensions</td>
</tr>
<tr>
<td>E</td>
<td>Economic dimension of sustainability</td>
<td>Concerns economic needs, infrastructure, distribution of wealth, control over resources, overconsumption, etc.</td>
</tr>
<tr>
<td></td>
<td>Environmental dimension of</td>
<td>Concerns ecological needs, pollution, climate change, waste management, green energy, etc.</td>
</tr>
<tr>
<td><strong>sustainability</strong></td>
<td><strong>Evaluand</strong></td>
<td>Something that is being evaluated, object under evaluation (e.g., products, policies, programs)</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Evaluation</strong></td>
<td>The systematic process of determining the merit (quality), worth (value), and/or significance (importance) of evaluands (e.g., programs, policies, and products) or evaluees (e.g., personnel), or the product thereof. Professional evaluation involves the use of systematic investigation to collect and synthesize factual information (what so?) to render evaluative conclusions (so what?) about an evaluand’s goodness, value, and importance.</td>
<td></td>
</tr>
<tr>
<td><strong>External evaluation</strong></td>
<td>Evaluation conducted from outside an organization or program; the evaluator is not on the pay role of the organization that designed or implemented the program (see Davidson, 2005a)</td>
<td></td>
</tr>
<tr>
<td><strong>F</strong></td>
<td>Formative evaluation</td>
<td>Improvement-oriented determination of merit, worth, and/or significance with the intent to inform decision making about the state of an evaluand’s/evaluee’s (e.g., program, policy, personnel) components or dimensions. This type of evaluation supports decision making about which program components or staff member competencies require improvement.</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td>Global bar</td>
<td>Involves setting or requiring minimum levels of combined or aggregated performance across all dimensions or components, normally using numeric indices (see Coryn, 2006).</td>
</tr>
<tr>
<td></td>
<td>Goal-free evaluation</td>
<td>The determination of merit, worth, and/or significance without explicit consideration of a program’s stated goals or objectives. Goal-free evaluation considers what an evaluand is doing instead of what it intended to do. Needs assessments are central elements in goal-free evaluation.</td>
</tr>
<tr>
<td></td>
<td>Grading</td>
<td>Assigning evaluands to an ordered set of categories, with the order corresponding to a metric of merit</td>
</tr>
<tr>
<td><strong>H</strong></td>
<td>Holistic bar</td>
<td>Involves a visual inspection (i.e., non-numeric) of performance across all dimensions or components, where performance across all must meet a minimum in order to ‘pass’ (see Coryn, 2006).</td>
</tr>
<tr>
<td></td>
<td>Holistic evaluation</td>
<td>The whole evaluand is considered without separating parts</td>
</tr>
<tr>
<td></td>
<td>Human dimension of sustainability</td>
<td>Concerns basic human needs such as food, shelter, health, etc.</td>
</tr>
<tr>
<td><strong>I</strong></td>
<td>Impact</td>
<td>Intended, unintended, anticipated, and unanticipated effects on targeted and non-targeted populations; usually referring to long-term effects and outcomes (see Davidson, 2005a; Frechtling, 2007)</td>
</tr>
<tr>
<td></td>
<td>Impactee</td>
<td>Everyone who experiences change due to the evaluand, including individuals who are directly affected by an intervention (i.e., downstream direct impactees) and individuals and organizations that are NOT involved in the evaluand and</td>
</tr>
</tbody>
</table>
are NOT direct recipients of the evaluand, but are still impacted by the potential range of outcomes of the evaluand recipients (i.e., downstream indirect impactees), those directly involved in the program implementation (e.g., staff; i.e., mid-stream consumers), and funders, political supporters, etc. (i.e., upstream impactees), (see Scriven, 2006)

**Indicators**

Factors, variables, or observations that provide evidence for the performance on a given criterion. Sustainability indicators must be specific (relate to the criterion), measurable (or observable), usable (practical), sensitive (must readily change as circumstances change), available (data must be collectable); and cost-effective (see Bell and Morse, 2003)

**Input**

Material, non-material, monetary, and non-monetary resources of an evaluation object

**Internal evaluation**

Evaluation conducted from within an organization or program; includes self-evaluation and evaluation by peer from different programs or units within the same organization (see Caspari, 2004, p. 32); the evaluator is on the pay role of the organization that designed or implemented the program (see Davidson, 2005a)

**M**

**Merit**

Intrinsic quality; quality without consideration of cost

**Metaevaluation**

Formative or summative evaluation of evaluation processes and products. Standards against which evaluations can be assessed include, for example, the Joint Committee Standards for Program Evaluation, the American Evaluation Associations Guiding Principles for Program Evaluators, and the U.S. Government Accountability Office’s Government Auditing Standards (Yellow Book).

**Mnemonic device**

Memory aid

**N**

**Needs assessment**

A systematic approach for determining states of existence or levels of performance of people, programs, or organizations. The purpose is to set priorities, allocate resources, and/or determine evaluative criteria. In contrast to wants or ideals, needs are essential for people, programs, or organizations to exist and perform reasonably in a given context. When conducting needs assessments, it is important to consider (un)met and (un)conscious needs as well as performance and treatment needs.

**O**

**Omission**

Oversight, exclusion (Scriven, 1991; 2007)

**Opportunity cost**

Activities or services that could have been implemented if resources had been allocated differently; forgone alternatives (Davidson, 2005a; Mathison, 2005)

**Outcome**

Usually intended, but also unintended change occurring as a consequence of the evaluand’s activities, progress toward goals.
(Frechtling, 2007; Davidson, 2005)

<table>
<thead>
<tr>
<th>P</th>
<th>Performance standards</th>
<th>Performance standards are specific values applied to the general criteria. They clarify what comprises different degrees of ‘good,’ ‘valuable,’ and/or ‘important’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy</td>
<td>Written plans that are informed by evidence, change focused, inclusive, strategic, causal, realistic, flexible, and outcome oriented (see Owen, 2006, p. 26) with the intend to guide decision making and action in specified contexts (see Davidson, 2005, p. 244). Policies can exist on local, national, and international levels.</td>
<td></td>
</tr>
<tr>
<td>Process</td>
<td>That what is being implemented within the realm of a program, including consideration of inputs, activities, services, and outputs</td>
<td></td>
</tr>
<tr>
<td>Product</td>
<td>A concrete result of a performance, task, production, or other process including outputs, outcomes, and impacts</td>
<td></td>
</tr>
<tr>
<td>Profiling</td>
<td>To graphically exhibiting grades, not scores, on the relevant dimensions of merit, worth, and/or significance</td>
<td></td>
</tr>
<tr>
<td>Program</td>
<td>A set of planned activities or services intended to address a need or other goals for a specified target group (see Davidson, 2005a; Owen, 2006)</td>
<td></td>
</tr>
<tr>
<td>Project</td>
<td>See program; usually a small scale program that is more refined in terms of time and scope.</td>
<td></td>
</tr>
<tr>
<td>Project-oriented sustainability</td>
<td>The target group or project holder continues the evaluand in its own interest for its own purposes over time (after Stockmann in Caspari, 2004, p. 67)</td>
<td></td>
</tr>
<tr>
<td>Proposal</td>
<td>A written plan/offer that specifies a program or an evaluation, associated prices, terms and conditions, products, goals, etc.</td>
<td></td>
</tr>
<tr>
<td>Purpose/use-oriented sustainability</td>
<td>Other groups or project holders adapt the evaluand for their interests, purposes, and uses (after Stockmann in Caspari, 2004, p. 67)</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Ranking</td>
<td>An operation used to place evaluands or evaluates in an order of merit (worth or significance) on the basis of their relative performance on a measurement or observation.</td>
</tr>
<tr>
<td>Reach</td>
<td>The breath of impacts resulting from an evaluation object geographically and over time</td>
<td></td>
</tr>
<tr>
<td>Risk</td>
<td>Factors that affect or are likely to affect success of an intervention; also negative consequences of an intervention to human life, health, property, or the environment</td>
<td></td>
</tr>
<tr>
<td>Rubric</td>
<td>Description of the meaning of a level of performance (e.g., inadequately, adequately, or exceptionally; scales can vary)</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>Refers to replicating, expanding, or increasing program efforts to extend positive outcomes and impacts</td>
<td></td>
</tr>
<tr>
<td>Scoring</td>
<td>Involves assigning numeric quantities, usually in terms of performance, on which to represent merit</td>
<td></td>
</tr>
<tr>
<td>Significance</td>
<td>Importance, relevance</td>
<td></td>
</tr>
<tr>
<td>Social dimension of sustainability</td>
<td>Concerns social needs, ways of organization, governance, and human interaction, etc.</td>
<td></td>
</tr>
<tr>
<td>Stakeholder</td>
<td>Those with a stake or some sort of investment in the evaluation object (see Davidson, 2005a; Frechtling, 2007)</td>
<td></td>
</tr>
<tr>
<td>Summative evaluation</td>
<td>Accountability-oriented evaluation that seeks to determine the merit, worth, and/or significance of an evaluand in order to inform decision making about the evaluand. This type of evaluation aids decision making about whether to continue or terminate a program, or hire or fire a staff member.</td>
<td></td>
</tr>
<tr>
<td>Sustainability</td>
<td>The capacity to exist (e.g., projects, programs, mankind) or continue (e.g., human, social, economic, and/or environmental benefits).</td>
<td></td>
</tr>
<tr>
<td>Synthesis</td>
<td>The process of integrating a set of ratings or performances on several dimensions, components, or criteria into an evaluative conclusion</td>
<td></td>
</tr>
<tr>
<td>Systems-oriented sustainability</td>
<td>The evaluand is being implemented system-wide to enhance performance across the system (e.g., educational or health systems); (after Stockmann in Caspari, 2004, pp. 67-77)</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>Generally refers to the logic or guiding framework for a program; a set of assumptions about how a program works; also referred to as logframe, logic model, theory of change (see Frechtling, 2007)</td>
<td></td>
</tr>
<tr>
<td>Top-down evaluation</td>
<td>An elite group (e.g., the funder) is requesting the evaluation.</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>Criteria of goodness, worth, and importance</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>Assigning levels of importance to components or dimensions of an evaluand or evaluee to indicate their relative or absolute importance</td>
<td></td>
</tr>
<tr>
<td>Worth</td>
<td>Material and in-material value</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C. Expert Recruitment and Consent

Western Michigan University
Department of: Interdisciplinary Evaluation
Principal Investigator: Michael Scriven
Student Investigator: Daniela Schroeter

March 30, 2008

Dear Expert:

My name is Daniela Schroeter. I am a doctoral candidate in the Interdisciplinary Ph.D. in Evaluation Program at Western Michigan University where I am currently working on my dissertation, titled “Evaluation of sustainability for sustainability”. I am writing to invite you to participate in a validation study of my Sustainability Evaluation Checklist (SEC). The intent of the SEC is to provide a practical guide for practitioners in development contexts to help them cover all the important items they should consider under this concept.

You have been selectively pre-qualified to participate in this study based on your exemplary work in evaluation and/or areas related to sustainability and sustainable development. Your knowledge and expertise will be highly valuable in assuring that the checklist is complete, correct, and useful.

The checklist provides guidance for general evaluation tasks and criteria of specific relevance to sustainability. Specifically, it distinguishes between evaluation OF sustainability (i.e., How well an evaluation object is being sustained?) and evaluation FOR sustainability (i.e., How well does an evaluation object address the larger concerns faced within sustainable development?).

If you agree to participate, I will provide you with the checklist and arrange for the most convenient mode of participation: via e-mail, web-based questionnaire, or interview. If you prefer the interview, please provide me with a good time during which I can reach you within the next four weeks. You can choose any dates and times. I will respond to you with a copy of the checklist, guidelines for providing the feedback, and any additional information you may request.

In addition to the time required to read the checklist, the interview will take about 30 minutes.

Your responses will be treated confidentially. However, if you feel that it is important for this research to disclose your name, you can attach a written request via e-mail.
If you are interested in participating in the study and willing to volunteer some of your valuable time for providing critical feedback on the checklist, please reply to this e-mail with your preferred contact information.

As a thank you for your participation, you will receive a summary of key findings from the study as well as the revised and improved checklist.

This study has been exempt from Western Michigan University’s Institutional Review Board.

I look forward to hearing from you.

Sincerely,

Daniela C. Schroeter

PhD Candidate in Interdisciplinary Evaluation
Western Michigan University Evaluation Center
http://evaluation.wmich.edu/phd/
Associate Editor of the Journal of MultiDisciplinary Evaluation
http://evaluation.wmich.edu/jmde/
E-mail: Daniela.Schroeter@gmail.com OR Daniela.Schroeter@wmich.edu
Cell: (001)269-267-8227
Appendix D. Specific Request to Expert

Western Michigan University
Department of: Interdisciplinary Evaluation
Principal Investigator: Michael Scriven
Student Investigator: Daniela Schroeter

April XXX, 2008

Dear <expert>:

Thank you very much for agreeing to participate in my dissertation research project. Please find the sustainability evaluation checklist attached. Please read the checklist and think about the following questions:

1. What is missing from the checklist?
2. What components or checkpoints are not necessary?
3. Are there any other errors or problems that need to be addressed?
4. What, if anything, did you like about the checklist?
5. Do you have any suggestions for how to improve the checklist?

I would like to contact you within the next few weeks to discuss these questions with you. I will also ask you to provide some general information about your experience and involvement in evaluation. Please let me know what dates and times work best for you and I will arrange an appointment. The interview will take approximately 30 minutes.

Sincerely,

Daniela C. Schroeter

PhD Candidate in Interdisciplinary Evaluation
Western Michigan University Evaluation Center
http://evaluation.wmich.edu/phd/
Associate Editor of the Journal of MultiDisciplinary Evaluation
http://evaluation.wmich.edu/jmde/
E-mail: Daniela.Schroeter@gmail.com OR Daniela.Schroeter@wmich.edu
Cell: 269-267-8227
Appendix E. E-mail to Listservs

Western Michigan University
Department of: Interdisciplinary Evaluation
Principal Investigator: Michael Scriven
Student Investigator: Daniela Schroeter

Dear list members:

My name is Daniela Schroeter. I am a doctoral candidate in the Interdisciplinary Ph.D. in Evaluation Program at Western Michigan University. I am writing to invite you to participate in a validation study of my Sustainability Evaluation Checklist. I have been on your listserv for quite a while and feel that many of you may be interested in my study as it relates to sustainability evaluation in international contexts. Your knowledge and expertise will be highly valuable in assuring that the checklist is complete, correct, and useful.

The checklist has been developed based on a literature review and provides guidance for general evaluation tasks as well as criteria of specific relevance to sustainability. Specifically, it distinguishes between evaluation OF sustainability (i.e., How well an evaluation object is being sustained?) and evaluation FOR sustainability (i.e., How well does an evaluation object address the larger concerns faced within sustainable development?).

If you agree to participate, I will provide you with the checklist and a survey that asks questions about the checklist and about your professional background. In addition to the time required to read the checklist, the survey will take approximately 20 minutes of your time. Your responses will be treated confidentially. If you know of any person who may be interested in this project and who is not on this listserv, please forward this e-mail respectively.

If you are interested in participating in the study and willing to volunteer some of your valuable time for providing critical feedback on the checklist, please reply favorably to my personal e-mail address: Daniela.Schroeter@gmail.com. I will respond to you with a copy of the checklist and the survey. Please indicate if you prefer taking the survey via a web-based link or a word document in which you can save your answers.
If you decide to participate, you will receive a synthesis of the responses to the survey as well as the improved checklist for use in your organization.

If you have any questions or concerns, please do not hesitate to contact me.

Sincerely,

Daniela C. Schroeter

PhD Candidate in Interdisciplinary Evaluation
Western Michigan University Evaluation Center
http://evaluation.wmich.edu/phd/
Associate Editor of the Journal of MultiDisciplinary Evaluation
http://evaluation.wmich.edu/jmde/
E-mail: Daniela.Schroeter@gmail.com OR Daniela.Schroeter@wmich.edu
Cell: 269-267-8227
Appendix F. HSIRB Notice (Scan)

Date: March 24, 2008
To: Michael Serven, Principal Investigator
    Daniela Schrock, Student Investigator for dissertation
From: Amy Naugle, Ph.D, Chair
Re: Approval not needed for protocol 08-03-15

This letter will serve as confirmation that your project "Validation Study of the Sustainability Evaluation Checklist" has been reviewed by the Human Subjects Institutional Review Board (HSIRB). Based on that review, the HSIRB has determined that approval is not required for you to conduct this project because you are studying the checklist and not collecting personal information about individuals. Thank you for your concerns about protecting the rights and welfare of human subjects.

A copy of your protocol and a copy of this letter will be maintained in the HSIRB files.
Appendix G. Invitation to Survey

Western Michigan University
Department of: Interdisciplinary Evaluation
Principal Investigator: Michael Scriven
Student Investigator: Daniela Schroeter

Dear [FirstName] [LastName]:

Thank you very much for agreeing to participate in my dissertation research project. I hope you received and enjoy reading the checklist and have many ideas for improving its content and usefulness.

The questionnaire about the checklist can be accessed online via:
[SurveyLink]
This link is uniquely tied to this survey and your email address, please do not forward this message.

The survey includes some rating items and open-ended items about the validity and usefulness of the checklist. It also includes some questions about your experience and involvement in evaluation. The survey will take approximately 20 minutes.

If you have any problems, questions, or concerns, please do not hesitate contacting me. I can send you the survey via word or in any format you prefer.

The survey will be open until May 5, 2008. I will send weekly reminders.

Thank you for your participation!

Sincerely,

Daniela

Daniela C. Schroeter
PhD Candidate in Interdisciplinary Evaluation
Western Michigan University Evaluation Center
http://evaluation.wmich.edu/phd/
Associate Editor of the Journal of MultiDisciplinary Evaluation
http://evaluation.wmich.edu/jmde/
E-mail: Daniela.Schroeter@gmail.com OR Daniela.Schroeter@wmich.edu
Cell: (001)269-267-8227

Please note: If you do not wish to receive further emails from us, please click the link below, and you will be automatically removed from our mailing list.

[RemoveLink]
Appendix H. Interview Protocol

<table>
<thead>
<tr>
<th>R/E(^{79})</th>
<th>Key questions</th>
<th>Potential follow-up(^{80})</th>
<th>AET(^{81}) (in min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Hi, this is Daniela speaking. Thank you so much for taking the time to contribute to my dissertation research.</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>May I begin by asking you a few questions about your experience with evaluating sustainability?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>3</td>
</tr>
<tr>
<td>R</td>
<td>Did anything occur to you as missing from the checklist?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>6 – 8</td>
</tr>
<tr>
<td>R</td>
<td>What components or checkpoints are not necessary?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>9 – 11</td>
</tr>
<tr>
<td>R</td>
<td>Are there any other errors that need to be addressed?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>12 – 14</td>
</tr>
<tr>
<td>R</td>
<td>What other problems do you perceive?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>15 – 17</td>
</tr>
<tr>
<td>R</td>
<td>What, if anything, did you like about the checklist?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>18 – 20</td>
</tr>
<tr>
<td>R</td>
<td>Do you have any other suggestions for how to improve the checklist?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>21 – 23</td>
</tr>
<tr>
<td>R</td>
<td>Thank you very much for your responses and the great advice. May I ask you a few additional questions?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>24 – 26</td>
</tr>
<tr>
<td>R</td>
<td>Do you have a preferred definition of “sustainability” or items that should be included?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>...</td>
<td>x</td>
<td>25 – 27</td>
</tr>
<tr>
<td>R</td>
<td>How do you understand the term “evaluation”?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

\(^{79}\) R = Researcher (Daniela Schroeter); E = Expert

\(^{80}\) Emerging follow-up questions will be noted for later interviews, if questions can contribute to further understanding about the strengths and weaknesses.

\(^{81}\) AET = Accumulative estimated time
<table>
<thead>
<tr>
<th>E</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>What do you perceive as the key challenges in the area of evaluating sustainability?</td>
</tr>
<tr>
<td>E</td>
<td>...</td>
</tr>
<tr>
<td>R</td>
<td>Is there anything else you would like to share or that I forgot asking?</td>
</tr>
<tr>
<td>E</td>
<td>...</td>
</tr>
</tbody>
</table>

Again, thank you very much for your time. Your responses are very helpful. I will be sending you a synthesis of the responses from all experts as well as a revised version of the checklist. Have a wonderful week(end)!

<table>
<thead>
<tr>
<th>Key questions</th>
<th>Potential follow-up (in min)</th>
<th>AET (in min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E ...</td>
<td>x 28 – 30</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>x 31 – 33</td>
<td></td>
</tr>
<tr>
<td>E ...</td>
<td>x 34 – 36</td>
<td></td>
</tr>
</tbody>
</table>

Total: ~ 30
Appendix I. Online Questionnaire

The sustainability evaluation checklist (SEC)

Please indicate your level of agreement with the following statements about the checklist.

1. The SEC is:

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Not at all</th>
<th>Slightly</th>
<th>Somewhat</th>
<th>Very</th>
<th>Completely</th>
<th>Don’t know/understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptable to differing cultural contexts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensive (i.e., complete)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coherent (i.e., items do not contradict each other)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concise (i.e., to the point)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concrete (i.e., tangible)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasible (i.e., viable)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy to follow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy to implement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important (i.e., valuable)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevant (i.e., related to the field)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useful (i.e., practical, helpful)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid (i.e., logically correct, legitimate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other. Please list another criterion for the checklist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. The checklist is useful for ...

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Not at all</th>
<th>Slightly</th>
<th>Somewhat</th>
<th>Very</th>
<th>Completely</th>
<th>Don’t know/understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Considering all aspects of sustainability evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Considering ethics in sustainability evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Considering those in need</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determining cost-effective ways for evaluating sustainability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developing an appropriate evaluation methodology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluating sustainability evaluations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifying criteria of specific relevance to sustainability evaluations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Using the SEC would ...</th>
<th>Not at all</th>
<th>Slightly</th>
<th>Somewhat</th>
<th>Very</th>
<th>Completely</th>
<th>Don’t know/understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change the way in which sustainability evaluation is conducted.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improve sustainability evaluation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other. Please list another consequence of using the checklist below</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. **Using the SEC would ...**

4. **Please list key WEAKNESSES of the checklist:**

   4a. What is missing?

   4b. What components or checkpoints are not necessary?

   4c. What areas are confusing?

5. **Are there any other errors or problems that need to be addressed?**

6. **What, if any, are the STRENGTHS of the checklist?**

7. **Do you have any suggestions for how to improve the checklist?**
About you

Please tell me a little bit about yourself.

8. How many years have you worked in monitoring and/or evaluation? Please provide a whole number.

9. What is your role in monitoring and/or evaluation efforts? Select ALL that apply.
   - Evaluation funder
   - Internal evaluator
   - External evaluator
   - Program manager
   - Other, please explain

10. In what regions do you work? Select ALL that apply.
    - Australia/New Zealand
    - Caribbean
    - Central America
    - Eastern Africa
    - Eastern Asia
    - Eastern Europe
    - Melanesia
    - Micronesia
    - Middle Africa
    - Northern Africa
    - North America
    - Northern Europe
    - Polynesia
    - South America
    - Southern Africa
    - Southern Europe
    - South-central Asia
    - South-eastern Asia
    - Western Africa
    - Western Asia
    - Western Europe

11. What country do you call HOME?
12. What is your primary organizational affiliation? Select ONE.
   - Education
   - Profit
   - National non-profit
   - International non-profit
   - Government
   - Multi-lateral
   - Other, please explain

13. In what sectors do you conduct evaluations? Select ALL that apply.
   - Agriculture
   - Manufacturing
   - Infrastructure
   - Social services
   - Education
   - Trade/Commerce
   - Tourism/Travel
   - Energy
   - Food/Beverages
   - Financial
   - Telecommunications
   - Transport
   - Other, please specify

14. Have you received formal training concerned with monitoring and/or evaluation specifically (NOT statistics or measurement)?
   - Yes
   - No

15. If yes, what type of training did you receive? Select all that apply.
   - University program
   - On the job training (e.g., professional development)
   - Off the job training (e.g., a certificate program)
   - Other, please specify:

16. Do you usually consider sustainability in your evaluations?
   - Yes
   - No
17. If yes, please describe briefly, how you usually evaluate sustainability.

18. What is the HIGHEST educational degree you have obtained?
   - Primary education
   - Secondary school (e.g., high school)
   - Professional training (e.g., certificate, apprenticeship)
   - Undergraduate degree (e.g., baccalaureate)
   - Graduate degree (e.g., MA, MBA)
   - Postgraduate degree (doctorate)
   - Other, please specify

19. Please indicate your primary area(s) of expertise or discipline (e.g., sociology, rural development, etc.).

20. Do you have experience working with evaluation checklists?
   - Yes
   - No
   - If yes, please indicate the checklist you have used.

21. Is there any other information that you would like to share?
Appendix K. Interview Summaries

Interviews were summarized based on recordings. Recordings were not transcribed because (1) they were conducted in both German and English, and (2) the information gathered was not intended for use in linguistic text analysis.

Interviewee 1 (April, 2008)

The first interview was conducted with a professor who advises the German Ministry of Environment on issues related to environmental sustainability evaluation. The interview lasted fifteen minutes.

After clarifying the purpose of the checklist, the respondent was walked through the key components of the checklist and concepts were clarified. The respondent thought that the project made sense and that there are not many comparable tools available. The respondent’s primary experience was related to sustainability evaluation of environmental plans. For example, he was involved in the evaluation of the German Sustainability Strategy. Recent triple bottom line approaches to evaluating sustainability would make the evaluation of sustainability more complex. The respondent pointed out that ‘environment is environment’ and not sustainability.

The respondent requested additional time to read the checklist and provided further feedback via e-mail as well as via the paper version of the Web-based questionnaire. In the e-mail feedback, the respondent provided a more in-depth assessment of the checklist. The checklist would be (a) well differentiated; (b) provide a good, almost complete overview of the most important aspects of the method; (c) good for
describing programs in breadth and for their classification; and (d) as complex as the subject matter. However, there would be a few problems with the checklist: Its transferability appeared difficult. Terms, though explained in the glossary, would be too extensively sociologists' slang. Furthermore, stakeholders could profit more from the checklist than decision makers for whom the categories appear too soft, at least in terms of environment (there are checklists for land/soil, water, air, toxicity, biodiversity, etc.).

The following question should be better answerable: Are the most important problems named, and are the chosen measures leading to goal achievement? In essence, users have to move too much on a dissertation's level, although simplicity and practicality are asked for.

_Interviewee 2 (April, 2008)_

This interview was conducted in German and lasted for forty-three minutes. After explaining the larger context of the checklist and its user group, details of the study were explained. The expert suggested that he did not have enough time to study the checklist in detail prior to the interview, but thought that the checklist was very extensive.

_Users_

It would be important to consider the important users. Once the potential users were clarified, the respondent suggested the existence of a dilemma, that is, the users could be overcharged. On the one hand, the list must be detailed so that the user can make
practical use of it. On the other hand, users could be discouraged by its length and complexity. But if something small and handy would be delivered, there would be no room for explanation or much detail. This is a problem that could be further explored in the pretest (i.e., the questionnaire), where people could express whether they could make use of the SEC or not.

The issue of complexity

Complexity would be another problem associated with the SEC. A solution may be to provide an overview in the beginning and later go into the operationalization and differentiation. The anxiety of having to work through a certain number of pages and a glossary could be reduced. The overview could be a big table where the central criteria are summarized. Those are then further defined in the second part.

Parts A & B of the SEC

The respondent found the general considerations (Part A) very important and necessary, as users need to know something about the foundations of evaluation and the evaluand. The respondent thought that that part is very good.

In Part B, however, the three overarching criteria (merit, worth, and significance) would require further explanation regarding why exactly those three criteria were chosen. An alternative, especially in the international development context would be the five DAC criteria: relevance, effectiveness, efficiency, impact, and sustainability, where sustainability is one criterion. But then there is an overlap in that significance or importance is a lot like relevance in the DAC. The expert stated that he is not
against the three criteria, especially if they are explained in terms of quality, value, and importance. But further explanation is required, because other criteria exist.

**Macro and micro levels**

The expert assumed that I would therefore be familiar with his conceptualizations of sustainability. That conceptualization has always been criticized for being on the program level or micro level where the macro-level criteria usually refer to the three pillars. Therefore, he now differentiates between micro and macro levels, but feels that they are completely differing things. On the basis of a differing concept of sustainability, it is hard to say whether something is missing.

After explaining how I intend to integrate both levels, the respondent expressed agreement, though his concept suggests the existence of a transmitter. If the funding cycle is ended, the question is what an organization does with what was funded initially. Do they continue? Continuation would be central for sustainability, and there would be agreement about it. But continuation would not be sufficient, and discussions have moved beyond that which the responded tried to capture in his most recent book.

Continuation (of activities) is not enough because the external conditions change with the removal of funding. There may be innovation, and, for example, curricula may continue to exist for twenty to thirty years, and nothing changed and people are proud of it. It is sustained, but what has to happen, to rewrite curricula, to adapt, to improve, that did not take place? So it makes sense to look at evaluation of sustainability on the micro level to lead to evaluation for sustainability on the macro level.
Institutionalization

Programs often have to be institutionalized, because they need the bureaucratic processes, but criteria have to be changed over time according to changes in the environment. Innovation has to be integrated in programming and in bureaucracies. Bureaucracies may be a little slower than businesses, but they should be able to change. As you can see in Germany, service organizations (like the universities) have changed a lot within the frame of new public management. The question is whether organizations integrate mechanisms that allow for adaptation to changes in the environment.

About the evaluand

Within the description of the evaluation object, unintended impacts may not be considered sufficiently. The expert tries to assess impacts logically on three levels—structures, processes, and behavior, where sustainability is a long-term consideration. He cannot think of anything that can change. So he questions whether unintended impacts as a central element are sufficiently considered in this section.

General procedures in evaluation

The expert points out that the general procedures (Part A, Section 3) include the identification of criteria that are further elaborated in Part B of the checklist. He thought that the structure might be problematic, because people cannot work on this before they have read Part B. There is an analytic logic and a process flow. The analytic logic presented in the checklist is not necessarily good for process flow. The analytic scheme is good for the dissertation, but the process flow might be better
for use in the field. For that purpose, the user may have to leave the analytic logic, something that is logical based on the flow of the checklist. In the end, it could be argued that the checklist is a model that needs to be adapted to each case, but a process flow would be useful. You should consider this and what the advantages and disadvantages are. Maybe it is good to leave it as it is or to create the process flow. Maybe feedback from practitioners can tell you if people have problems with the flow.

Utility in international development

The respondent thought that the checklist could be useful, but not quite in its current format. His experience suggests that it may not get the acceptance it may need. Many development agencies are pessimistic about tools like this, although they may be quite useful. The problem is that there is a lot of evaluation, but little use of evaluation. The good thing about the SEC would be that it could be used ex ante and that would be very positive, because programs can still be changed and findings from evaluation can be used more widely. Of course it is important to begin using the checklist with planning. The problem is that it is not too user-friendly. The idea is great and should be supported, so there needs to be a format that makes people look forward to use it: Provide a one-page overview in the beginning, —and then you can get it thinner by putting all the subpoints into a separate part of the document. At least it can look more user-friendly. But then you have to move forward and backward.
References

References should not be included in the checklist, but a reference to an academic discussion about who influenced the checklist is important. But in the checklist, this would be bothersome as it is supposed to be an operative instrument. Adding a reference list at the end is a good alternative.

Interviewee 3 (April, 2008)

This interview was conducted as a follow-up to a survey response in which the respondent provided a phone number to clarify questions. This interviewee did not belong to the original expert sample.

The expert also provided extensive feedback via the Web-based survey. One of the most critical points was that the expert perceived a lack of guidance for formulating good questions on sustainability evaluation. The question is, what should a person get out of this? The expert thought people should get achieve an appreciation for sustainability evaluation and its complexities, an understanding if they can or cannot do it themselves and should hire someone.

One of the key questions is, who is the audience of the tool? The expert’s perception was that the audience would not be evaluation specialists and thought that the tool would develop introductory competence to evaluate the comprehensiveness of the evaluation strategies used. It would not specifically address sustainability ultimately, but provide consumers of evaluation with what they should look for.
Another question is, what the checklist is useful for? It is minimal for evaluating sustainability. It is useful for prompting discussion about sustainability.

It would be good that the checklist takes on the full parameter from teaching evaluation 101 in terms of a very difficult topic. Most of the development interventions do not continue in any way related to how they were funded originally. The risk criterion, for example, is inadequate because it is too brief. Each criterion must be further expanded on.

*Interviewee 4 (April, 2008)*

The interview was conducted face-to-face and lasted fifty minutes. This expert is a highly published individual in evaluation with extensive experience in the nonprofit sector. The interviewee also responded to the questionnaire.

**Experience with evaluating sustainability**

The expert directed a foundation’s evaluation unit where a key question pertained to the sustainability of funded projects beyond the foundation’s funding. The continuation of the project was the primary concern. Economic issues were one factor asked about. In essence, whether grants received continued support or further support. They looked at internal and external forms of support as well as personnel turnover and its effects on sustainability.

**Evolutionary sustainability**

One of the interesting things that were found related to the evolution of the original intent of the grant. Often there was a specific purpose of the grant, but once the
funding was terminated and continuous support was provided internally, the emphasis of the grant was changed. For example, the target population or the emphasis of the grant was narrowed down or expanded, depending on the resources available. Maybe based on learning, procedures were changed. It was found that the original intent continued, but not in the same form. So, the question was, at what point is this evaluand a different evaluand, more specifically, at what point is it so different that one could say that the original intent is discontinued?

Therefore, the evolution of projects became one of the focal points in this foundation’s sustainability studies. One question was, how different was a project after one, two, five, and ten years? In each case, the evolution of the project was observed to make subjective determinations as to whether and at what point the project was no longer original, that it was so different that it had to be called something different. Other things of interest were factors pertaining to why or why not a project would be continued.

Users

The expert felt that the checklist would be for professional evaluators. A manager’s eyes would “glaze over it,” so he assumed that the purpose is to guide evaluators. The vocabulary is good for professional evaluators, but not for managers or developers. For them it is not clear enough and too technical.
The SEC: Part A

Integrity and ethicality are problematic concepts used in the checklist. The expert thinks they have to be included in the glossary if the terms are distinguished conceptually.

More clarity would also be necessary on the metaevaluation checkpoint. There may be confusion in using “i.e.” versus “e.g.” Validity and accuracy, they are about the same. When considering the PES, validity is only one standard of several accuracy standards, yet the most important one.

In the key concepts in the beginning of the checklist, evaluation should be distinguished from research for people with differing contexts. You could say, “Evaluation is different from research in terms of quality, value, importance versus generation of new knowledge.” This could be important in preventing people of thinking right away “experimental design.”

In determining the purpose of evaluation, knowledge generation should include the factors that affect sustainability so we learn why something is sustainable or not, the factors contributing to or hindering sustainability.

In “determine the direction of evaluation” there may be a third group of people, the investigators who are just interested in exploring sustainability without funding. But that are researchers, which would require differences between research and evaluation in the beginning.

In “define the timeframe under evaluation,” sometimes the evaluand vanishes quickly and sometimes it evolves very quickly. Furthermore there are cases where the
evaluand is supposed to complete quickly. The question is again, at what point is an evaluand a different species. It is a philosophical and a methodological question. You don’t want to continue to collect data on an entity that no longer has the identifying characteristics or the original characteristics of the evaluand. The perspective is important in that the a funded evaluation reports back to a client who is interested in original intent, while grassroots efforts in this case present research that informs process.

Under “collect data” in section 3, “Collect data on the quality of the instruments especially reliability and validity” should be added.

The SEC: Part B

The list of criteria is good, though not necessarily exhaustive. The experts said he wished he had this checklist 10-15 years ago.

The glossary

“Standard” should be added as it is an important term in your checklist. It requires a place for people to go to.

Summary

The checklist is very useful and provocative. It stimulates thinking. The distinction between evaluations of/for sustainability was not that important to the expert as the difference between evaluation and research and being clear about the need to look at why something is sustainable and how it is sustainable and then the evolution questions. The experts felt that both evaluation of and for sustainability would be important.
The interview was conducted face to face and lasted about 60 minutes. This interviewee is an international development expert with many years of experience. In the beginning of the interview, discussion focused on the type of projects the interviewee is engaged in internationally, that is, agricultural development projects. Experiences in the field and in differing countries were exchanged prior to focusing on the SEC. At the end of the interview, the expert also completed the questionnaire and provided a copy of his thought process when applying the checklist.

The SEC

The experts said that he read through the checklist with a proposal in mind, so he read it in an applied way, “testing it.” He thought that it was a good exercise that helped him to separate dimensions and separating merit, worth, and significance. These would bring out very different things. How much does it change the yield and the economic value, but there are other merit issues and competing things, such as seeds versus fertilizers to increase production, but soon you are dependent on a distribution chain. With oxen, people are self-sufficient, can work on their own and it does not matter where politics goes. For example, last year there was a country where there were elections issues resulting that trucks were stopped and seeds could not be distributed. People with oxen are independent of such issues.
Sustainability

In international development, sustainability has two main components he thinks:
First, an intervention has to be economically viable for a set of people. Second, will it
not do any harm to the environment? Any use of energy can either do harm or good.
For example, we recommend tools and practices but whether they accept it and use it
the ways we recommend, are open questions that depend on cultural practices and
levels of understanding. There are many results of given interventions, for example,
the yield may grow exponentially and suddenly a man wonders why he does not have
enough wives to weed it all. If that is the case the yields per unit diminish and
educational learning things have to take place, which are more difficult. Several
issues like this may occur due to interactions between environmental, cultural, and
technological issues. For examples, certain animals may not do well in certain
climates. In some cases, vaccinations are available to fix the problems in others not.
The objective of our international development organization is to pull together the
knowledge base of our forefathers to help people.

Self-reliance

The self-reliance dimension in the SEC was not as strongly addressed as it should be.
It is an important issue. We have to be interdependent to be successful, but in those
field situations, this interdependency may not be that critical and we have to be
careful to not make people too dependent. It is a dimension we have to remember. It
is a basic security issues. This is one of the reasons why it is harder to convince a
traditional farmer to change, because there is no safety net other than what the
grandfather or the experience of the family is.

**Connections between self-reliance and solidarity**

He is not advocating against solidarity, but in a family unit it's a mothers concern. It
would be nice if people would specialize. The farmer could trade food, the blacksmith
help with fixing tools, etc. But usually people want to do it all themselves. In certain
seasons, everyone is in the fields. So, if the farmer needed a tool fixed and would go
to the blacksmith, he would not find him, because the blacksmith is working in his
fields as well.

**Commissions**

The expert did not think so. He agreed it is dense and first screened the list and then
went to get a case he could apply it on and he felt it was a good exercise.

**Simplicity/complexity**

When asked if the checklist should be simplified, he referred to one of the evaluators
in the community who visited him and took a quick look at the checklist. That
evaluator said: “So quality is now merit.” Generally, the expert felt that terminology
was explained well enough, but the language could be simplified for people
internationally.

Practicality could be improved, but he sees the streamline benefit of a checklist. He
referred to a checklist he had developed years ago to simplify legal processes for
nonprofit organizations and thought that a checklist is a good idea and beneficial to
practitioners. He feels that the checklist should not be much expanded. The use of the
Internet would be a good idea to interlink terminology with the glossary. His organization works with CDs and DVDs that can be run over a computer and are less expensive than books.

In terms of the terminology, words could be exchanged to simpler ones, but that would not solve the problem, just change it. For example, instead of “evaluand” it could say “object.” However, both are very abstract.

To really simplify it, the checklist would have to be specified for sectors: agriculture, more industrial programs, etc., but then it would not have wide applicability and utility.

Application

The expert said that he tried to apply it, and it seemed to work. He said that he has a graduate degree in sociology and is married to an evaluator, so he quickly made notes on different dimensions and specified the sets of criteria: merit, worth, and significance. He said, “What more can you expect than to stimulate the thought processes? You are not going to get the same product every time, because that person has a lot more baggage in their head before they even pick this thing up."

*Interviewee 6 (April, 2008)*

This telephone interview with an evaluation expert and practitioner lasted for seventy-five minutes. The expert also agreed to complete the questionnaire. The expert began with some “free associations” about the SEC and the issue of sustainability. His first thought was that a checklist would imply a list. That is, the
more things you can check off, the higher the score. That means it is a linear thought process or mode. He felt that this was not true for the SEC.

He could imagine sustainability happening for a lot of reasons. A program may or may not meet the needs of the stakeholders, but is a line item in the budget, thus exists either way. The program is highly sustainable, good or not good. So programs can be sustained for many different reasons and many different combinations of factors. On a formal level, there could be a lot of supporting cross-linkages. On the other hand, there could be few cross-linkages but very strong cultural beliefs, so people may think it is inappropriate not to have the program. These programs could be equally sustainable. When scoring each on the values item and the cross-linkages item and adding both dimensions, it could make one program look more sustainable than the other, although this may not be the case, and that would be problematic. As the SEC does not treat dimensions diagnostically, the term “checklist” may have to be reconsidered.

Some programs could be less sustainable but more adaptable. Others could be hard to remove, but once they become brittle, go away fast. So the question of defining sustainability relates to the question of adaptability versus brittleness. From an evolutionary biology perspective, sustainability has a lot to do with its context. So if you have a rather weak program that lives on a very gentle fitness landscape, it may be very sustainable, while a more intensive landscape may destroy it fast. So there is a contextual issue. This can be seen in healthcare (e.g., polio) but also in education (e.g., public schools in comparison to charter schools). The question whether
something is sustainable really has to do with external context. It may be resistant or very brittle.

The checklist idea does not work for him for the reason that it cannot be a diagnostic tool. It needs to be more of a model, a structural and/or behavioral model related to linkages and support mechanisms.

Some programs are designed to not be sustainable. It is a relationship of the supportive infrastructures, which may not fit with the checklist.

There is an assumption that the program has to be good to be sustained, though lousy programs could be very sustainable. It is very value-laden to say a program has merit and depends much on the politics. I may believe that a program has no merit and find it not worthwhile to be sustained, but the politics are on the side of the program, so it’s a tricky issue.

The notion sustainability is tightly linked to aspects of evaluation in the SEC. The question of whether a program is sustainable is not a matter of its goodness or value conceptually. On an intellectual level, the issue of something being sustainable can be assessed independently from the actual merit or worth of a program. As indicated by these notions, the expert understands sustainability as continuation only and independent of the notion sustainable development.

The expert also stated that one should not try to make programs sustainable and that whether a program is sustainable is a separate question.

When reminded that the SEC is interested in bridging the gap between program sustainability in general and in light of broader issues of sustainable development, the
expert suggested the emergence of a logic model. If there is a goal of sustainable
development, then there would be a combination of transitory and permanent effects.
Then a transitory program should go away, while others would stick around for a long
time. For example, if international development interventions are intended to
transition from wood burning to coal burning, then one would have to say that it is a
transitory intervention. While coal burning might be better than cutting down all the
woods, we know that it is not environmentally sound either, so we want it to go away.
Next, the question of whether there are programs that are not related to sustainable
development was addressed. For example, recent foreclosure programs are hardly
linked to sustainable development. Nor are programs designed to help keep military
families together to assure soldiers go to Iraq. This may be an important program, but
does not relate necessary to sustainable development.
The expert said he didn't know how to improve the checklist, but that the relationship
between sustainable development and sustaining programs would be problematic. It
may be better to come up with a few different models for sustainability. For example,
in a general fitness landscape you would want a lot of coherence in your organization.
But in the jagged fitness landscape you want a lot of variability to support
adaptability. The issue of fast-moving and slow-moving environments is important.
There are different aspects of sustainability: supporting infrastructure, being part of
an organizational structure, being a line item in a budget. The infrastructure has to do
with the environment and cross-linkages; the second part has to do with
organizational culture. These are practical issues that can help you figure out plans for sustainability.

You could look at it from a structural point of view, from a cultural point of view, from an environmental adaptive point of view, and pick different strategies to go about looking at it and make suggestions.

If it was possible to articulate principles such as would a change in this program upset many other strong programs? Imagine two different programs—one has a lot of cross-linkages (e.g., community development). You could look at the program and describe the nature as supportive cross linkages are critical for its success. Another program may be about diabetics that have to change their life. Such issues are based on belief systems and cultural issues, so cross-linkages do not even make sense. Sustainability in this context relates to beliefs and maybe family support systems rather than organizational linkages. We try to incubate this belief into people. This is not an issue of line items or organizational linkages, but politics or religious groups. It’s a different set of issues. So with programs like this, you may have to think about economic and environmental perspectives; and for other programs, you need to consider the social and cultural dimensions more specifically. So it is not clear-cut, it is never pure, but if you can take a broad cut, you can distinguish by types of programs. That is one big distinction you can make.

The link between sustainability and sustainable development may muddy the water. A lot of the points in the checklist make sense. For example, he really liked the notion of policy, program, and project as differing yet integrated types of evaluands. Projects
and programs involve people and organizational politics, etc., while policies are on paper and have a lot to do with bureaucracy and consensus-building. Another thing may be to consider differing stakeholder groups. Not all stakeholders are equal. Reach relates to sustainability a lot and relates to linkages of a program to other places organizations. He thinks the sustainability and sustainable development should be considered separately. Ultimately you need to be in the edge of chaos stuff that has to do with adaptability and organizing fitness landscapes.

The criteria are tricky. He does not think merit and worth are the right things to think about in sustainability, but it is not so straightforward as the military example shows. The notion of synthesis is really important, not necessarily in relation to merit, worth, and significance, but sustainability is multidimensional, and thus requires a lot of integration.

Infrastructure makes sense, though responsiveness is problematic, because it is related to the environment/context, but knowing what change is relevant or not is hard to know. Include some items that deal with decision making regarding the context. The M&E checkpoint may be problematic similar to responsiveness. You just don’t necessarily know what is important or not. You do not know what you observe or do not observe. Considering unintended outcomes and impacts makes a lot of sense as well as the inclusion of opportunity cost, but global costs may not be easy to really to consider.
Separating criteria for evaluation of/for sustainability may be important. From a practical point of view, if people have tried it and it helps clients think about their program, the checklist is great.

*Interviewee 7 (April, 2008)*

This telephone interview lasted twenty minutes and was conducted with an international sustainable development expert. The expert also provided extensive comments via the questionnaire and direct comments on the checklist.

He thought the SEC is “very, very interesting and useful” and “very, very good.” The first part is well structured, but he got a little lost in the second part. The logic frame did not quite fit. That does not mean it is logic, but somehow it was tricky. All the criteria would be very relevant, but it misses a red line to keep it together. Sometimes the criteria of merit, worth, and significance are useful, but in terms of sustainability they may be a little misleading because it is hard to put everything in to these boxes. The second part requires a look back at Bellagio. Maybe that could be used to form the red line.

For example in section 4, the main points are good, but their specification seems a little redundant. The structure needs to be a little bit simplified. He said he got a little lost when trying to comment on the list.

The difference between ex ante and ex post evaluation as well as between monitoring and evaluation and metaevaluation must be clarified in the SEC.
He did not think anything would be missing. Learning processes could be made more explicit. In section 1, there is a point to determine the types of evaluation: holistic versus analytic. He thought that theory-driven evaluation should be explicitly named. The links between the evaluator and the audience are very important as well as and can change a lot depending on the role of the evaluator in the relationship with the stakeholders.

In section 2, stakeholders and impactees, it may be good to distinguish between internal and external stakeholders. Issues of scale and time: Time is especially important because the evaluation changes dependent on the time at which something is being evaluated. The same accounts for scale, where, for example, the increase in use of biofuels reduces emissions here, but increases deforestation in Brazil and increases the cost of maize and basic foods in Mexico. The client who is asking for the evaluation may define the timescale as well as the boundaries for the evaluation. An ethical dimension should be integrated more explicitly to overcome, for example, the strings that a client may put on a client. There are also a lot of political issues. The checklist is really useful. It helps to clarify critical issues systematically.

_Interviewee 8 (April, 2008)_

This telephone interview was conducted with an international development and evaluation expert. The conversation lasted forty minutes.

The expert thought that it is a useful topic that does not get much attention, but is really important. However, a lot of the SEC would not relate to sustainability, and it
would take a lot of time before getting into the sustainability issues. He understands that these initial concerns are important, but considering the amount of time available to program managers, it may take too long to get to the sustainability issues.

**General thoughts**

His experience is that virtually nobody cares about sustainability in practice. Most implementing agencies are concerned with implementing the project. Once the project closes, there is little interest from the side of donors or the governments in thinking about sustainability. So, some questions could be included to explore the donors’ interest in sustainability and how they might use the information. By the time you get around to sustainability, the program may be merged in organizational structures or just be abandoned.

A half-serious joke from Bangladesh: The brightest people negotiate with donors to get a new program. The next brightest people are responsible for implementing the project when loan disbursements are still pending. The dumb people are the ones put in charge of management once donor funding has ended.

This is an observation of the priority of sustainability the same as people care about impact assessment. Although it is really important, too little attention would be paid.

It is not even clear who would read an impact evaluation

He wondered if there should be more guidance on the methodology for evaluating sustainability. It is a difficult issue for a number of reasons: Sometimes it is prospective and the methodologies for that are difficult. Also how to deal with issues of attribution, how to deal with changes. So the question of the methodology may
need some guidance and whether you want to support people in using qualitative methods and help with economic analysis.

A lot of people talk about sustainability analysis, but that is just prospective analysis where people make heroic assumptions, but then when you look at earlier projects you find no support for the assumption made. Sustainability analysis is often just an arithmetic exercise that is based on tenuous grounds and heroic assumptions. It would be important if there should just be prospective analysis or if it is necessary to go back. In the prospective analysis it may be very useful to employ program theory to spell out the relationship between indicators, so after a few years in the program you can reflect back on the logic model and look at the assumptions you made and see if these things are on track or not. Program theory could be helpful. It is a learning tool that may be used prospectively. The main point is to have a conceptual framework for thinking about whether things are on track.

The respondent developed a sustainability checklist together with a colleague that was used a few times. That is government officials were sent to projects to assess project sustainability in agricultural extension. It was a checklist with a five-point scale and included five indicators under each of four dimensions: continued delivery, maintenance of infrastructure, institutional capacity, and support from stakeholders. The experts kindly sent the checklist to me after the interview. He found that a team of a few people could get a fairly good shot at it in a few days. It was something that was to be used in the field easily, and you could just hire someone from a university to have a look at those things. It gives you a fairly good idea of how the situation is.
Strengths

He thought it is something really useful, something that needs to be done. He thought it was good that included social and cultural dimensions. The checklists covered a lot of the important issues.

Checklist: A misnomer?

After explaining how different people may perceive checklists, the expert concluded that people have different expectations, so it is important to spell out what exactly the tool is doing, help people think about things or rating things. He felt it was certainly a checklist. The SEC could be put in two forms: One to design evaluation and another one where people can check off things. Calling it something like a tool would make it possible to talk about methodology, etc. Many people do a short checklist, something easy to use, and then you could have the more detailed document with it. This may be done over a longer period of time.

Section 2

It may be important to describe how sustainability relates to different types of the evaluands. What does sustainability mean for different evaluands?

Section 4

Strong support without having to identify with given political or certain groups which may give you the kiss of death as soon as the leadership changes

Commissions

He did not think there were any. It is a practical question, but nothing really needs to be cut out. It is very interesting.
Community participation

Community participation may increase sustainability, but there is little empirical evidence that this is true. It is an interesting question that can be illustrated with water supply projects. There are studies where different levels of participation have been explored. But sustainability is often dependent on how beneficial an intervention is. In some cases people, do not want to go to meetings or participate because it interferes with daily routine. Some people claim that it is beneficial to involve women in certain types of projects, without empirical evidence for it. Norman Uphoff provides good examples in relation to participation and sustainability.

Interviewee 9 (May, 2008)

This phone interview took place with an evaluation expert with extensive experience in various sectors. The conversation lasted one hour. The expert also responded to the questionnaire.

Sustainability is a “messy” area. The checklist’s take on it comes from yet another direction. He intended to test it but did not.

Provide an example of a community project that received funding for three years with one of the goals being that the program be sustained. Sustainability is an idea that can be looked at from different perspectives: (i) sustainability of activities where the activity is what is meant by sustainability from the beginning on; (ii) sustainability of the idea, where sustainability is represented by repeating events though the activity may be different across repetitions (the DNA of the activity is changed); (iii)
sustainability of capacity (human capacity, social capacity, infrastructural capacity, organizational capacity—all of which are located within community practice); and (iv) sustainability of results (may or may not be independent of the activity).

The integration of materials from outside the United States is noteworthy. In the U.S., it is very much about routinization, institutionalization, and standardization. From a community work and systems background, it is clear that sustainability relates to ambiguity of goals, which is highly relevant for environmental sustainability, and ambiguity of effect, when you have no idea what impact you are having. The big international development organizations spent millions to force methods onto projects to give them information about whether they have impact, and the answer is they do not. There is unreliability and turbulence and complexity and unknowingness about the environment, and that is highly relevant in less centralized countries where several layers of government activity exist. There are all these power structures and shifts, so you may be successful until the government changes and then there is no more support.

Differences between a checklist and a heuristic

A heuristic is a bit like a checklist, but is something that makes you think. A particular heuristic is the critical systems heuristic which might be useful for looking at sustainability. In certain environments sustainability is all about innovation and in others it is all about routinization. Adaptability is between the two like a compromise, it is more incremental. Innovation is you dissolve the problem, you reinvent yourself.
The critical heuristics checklist is a means to decide what you talk about with a client, not something you talk through with a client. The SEC may be seen as a thinking tool. “Ability to adjust to changing contexts”—that checkpoint I could write one hundred items on. It has huge epistemological and ontological implications.

Projects without environment

In the U.S., people look at systems as surreal, and in Europe they regard them as perceptions, so you can run a project as if it were a closed system. You can look at it and say that in reality, it is really successful in its environment, but the people running it run it as if it were a closed system. You run it as the perception, which then becomes the reality.

The checklist is enormously rich, and on every point you will find some people that agree and other that disagree.

Bringing the systems perspectives to the SEC

Systems theory is a huge field, much larger than evaluation. The question to ask is what is looking at something from a systemic field all about?

To bring in the system perspective, there are three things that you need to make sure that you are addressing:

1. Interrelations: How does the SEC deal with interrelationships? The simple, the complicated, and the complex interrelationships that occur within a project, within its environment, within your concepts.

2. Perspectives: People look at things from different perspectives. The realist perspective is that there is an objective reality, but the way people react with
reality depends on perspective, so depending on a given perspective things may be treated differently, have different purposes. How we react on things depends on how we see them.

3. Boundaries: Especially relevant to evaluation and better addressed in systems thinking than in evaluation—a value system is a boundary, but the way in which value systems are explored in evaluation is limited. What is in and what is out? What is relevant and what is not? What is marginal and what is not? These are the important issues in projects with a systemic approach. Holism is contrary to being systemic, though people widely believe they are related. You cannot include everything; that would lead to madness. You would be treating the world as it would be a closed system (see Churchman, an ethicist).

The systems world largely has given up trying to define what a system is because it was too unsystemic. Sustainability may get in that position, with its multiple purposes, perspectives, etc., so it is almost undesirable trying to define what it is. The checklist seems to be independent or accommodating all definitions that there are. “Of” sustainability seems tricky. It has a summative characteristic to it. Take another look at the boundary issue: Sustainability is very dependent on who is in and who is out.

*Interviewee 10 (May, 2008)*

This last telephone interview was conducted with one of the most distinguished experts in sustainable development. The interview lasted forty minutes. The expert
also responded to the questionnaire but did not provide any qualitative data in lieu of the scheduled interview.

First, the expert elaborated on general issues regarding the SEC. He suggested that he was not the most favorable about the checklist because he does these evaluations frequently and has his biases as to what goes well and what are problems. The level is too high. The separation between part A and B was less clear, particularly the second part of part B, which seemed more process-related to financial issues and time issues that he found more relevant to part A, which is more process-oriented than outcome-oriented. Part B should distinguish process and outcome.

Another issue in every part is a classification of the questions. It was not on a common ground. Sometimes there is a switch in the focus of the questions within the heading. These are highlighted in the particular places later on.

Some points are very subjective, leaving the evaluator with uncertain choices, particular when you have instructions, such as “consider.” That is very unclear. There needs to be more clarity as how to use the checklist.

There should be use of common language. We all should use the same language. It is very academic.

Part A, p. 2: Type of evaluation, the distinction between holistic and analytic is not practical. It does not help beyond academic work.

The evaluand: There is a problem with the mix of stakeholders and impactees. Either you want to define the stakeholders or the impactees. These questions are different. I
would not mix the two. Separate them out. I suggest making another bullet, so it is not indefinite.

The "nature of the evaluand," particularly the "by whom" issue, is mixing different things and makes it more complicated. There need to be two sets of questions. What has/is to be sustained? Then the question is by whom. I would not ask the question. It is easier to define what is to be sustained than by whom. If you separate it, you can define who is in charge in a second question. But again, you may mix two things. Those who are in charge for sustaining things may not be the ones responsible for it to actually occur.

Section 3—a lot of general comments on criteria and standards. These can be internal or external criteria and standards, which is important. If they are internal, you have to create them. If they are external, you have to find them in different sources. He was not sure what the checklist intended to do, the first or the second. It has to be clear that both internal and external criteria and standards have to be considered. This way, local issues can be emphasized or the criteria and standards can be adapted to them.

In the underlying logic between sections 2 and 3, there is a shift. In section two, it seems to be about sustainability evaluation while section three shifts to performance evaluation.

There needs to be two different types of evaluation when you talk about performance evaluation or outcome evaluation. There needs to be more clarity across the checklist. The expert pointed to several areas where miscommunication occurred. For example
“collect data from relevant sources,” it must be “all” relevant sources to make sure that the user checks for “all” relevant sources.

**Part B**

The introduction to the criteria implies a value judgment, and you do not give the criteria as to what is “good” sustainability. Sustainability in itself is already value-laden. Either you believe all sustainability is good or it is a matter of degree. It matters again where the standards are from to decide whether it is good or not so good.

Leadership is very difficult to measure unless you find some proxies. If you want to have a good measure of it, I doubt you could find it. There are several other places, where the question is how do you measure it?

Infrastructure is important, but we have a shift in the classification. Once you speak about some technical needs, I am not sure why adequate waste treatment is important. What were you thinking about putting these under one heading? It does not relate to the general explanation after the checkpoint heading. After I explained the rationale, he emphasized that this mixes up issues. He generated a thought process that allowed me to reconsider some checkpoints in terms of moving them to different sections.

Understanding the community—How can you measure it? This point is clear and makes sense, but can you measure it.

There is confusion between acceptability and equity that needs to be reconsidered. Intergenerational equity is good to have in but it is not possible to measure as it is about potential. So the level of generality to measure potentials is impossible.
Use of evidence is not clear. You do not say what to do with the evidence. This should be rephrased.

It is a helpful checklist, but the purpose needs to be clearer. One more thing is related to criteria of significance, when talking about scope and scale, but “scale” is not quite used. I know what you want to say, but the description does not really address a scale issue.

It is an okay effort, but as a student you face the problem that it is a mix between an academic and a practitioner’s approach.
Appendix L. Sustainability Evaluation Checklist (Revised Version)

Introduction

* Indicates terms explained in the glossary

Purpose: Why use the SEC?

The Sustainability Evaluation Checklist (SEC) is intended for use in planning and designing project and program evaluations of sustainability for sustainability within development contexts. As a tool, the SEC is designed to help users not only remember certain tasks involved in evaluation but also consider a wide array of criteria of importance to sustainability evaluation. Thus, the checklist aims to:

(a) reduce errors of omission*

(b) increase evaluation usefulness

In addition to planning and designing sustainability evaluations, the SEC may also be used to:

(a) generate ideas and discussion on key issues in sustainability evaluation

(b) support proposal writing processes

(c) compare existing sustainability evaluations to determine whether all important aspects have been met

Intended users: Who should use the SEC?

The SEC is intended as a guide for individuals who are frequently involved in internal* or external* evaluations of development projects or programs with a special interest in sustainability. These users may include:

(a) Evaluators and researchers who provide evaluation services

(b) Program planners, funders, and managers with an interest in evaluation
(c) Program recipients/participants/users who start their own evaluations

(d) Others who have an interest in evaluation

Experienced evaluators may benefit from the comprehensive nature of the checklist and use it as a point of reference or generator of ideas. Novice evaluators or those with limited exposure to evaluation may find value in the instructional elements of the checklist.

Characteristics: What are the parts of the SEC?

The checklist is built on Scriven’s (1982) and Fournier’s (1995) logic of evaluation and consists of two major parts: (A) general considerations and (B) criteria* of merit*, worth*, and significance*.

The general considerations, Part A, are subdivided into three sections: (1) grounding the evaluation, (2) about the evaluand*, and (3) general procedures for evaluation. As suggested by the title of the section, this part includes aspects of relevance in any evaluation. But it is imbued with elements of specific relevance to sustainability evaluations. For example, section 1 intends to clarify different perspectives on sustainability evaluation dependent on the time at which sustainability is assessed; section 2 looks at dimensions OF and FOR sustainability; and section three provides guidelines on the general procedures. This third section also provides the connection to part (B).

In part (B), criteria for evaluating sustainability are distinguished by (4) importance, (5) merit, and (6) worth. These criteria have been developed based on an extensive literature review and feedback from sustainability evaluation experts and practitioners. Section 4 looks at aspects that help determine whether sustainability is relevant in a given situation. Section 5 considers those aspects that enlighten capacity FOR sustainability, and supports determining if and to what extent aspects continue after initial resources are reduced or removed. Section 6 looks specifically at costs.

Because the checklist is relatively generic, users may find value in the “heuristic” characteristics of the checklist. That means it stimulates thinking that may encourage discussion within organizations and among evaluators, clients, stakeholders, and impactees.
Key concepts: What terms need to be understood to use the SEC?

*Sustainability*

The capacity to exist (e.g., projects, programs, mankind) and/or continue (e.g., human, social, economic, and/or environmental benefits*).

For programs, this usually means existence (temporal durability) beyond termination of initial support via mechanisms that have been used to develop the program. In terms of continued benefits, it means that programming does not negatively impact human survival on earth. For example, attitudes or practice may be eliminated/eradicated to allow for sustainable development of societies. That means not everything should be sustained nor is intended to be sustained.

*Evaluation*

The systematic determination of merit (quality), worth (value), and significance (importance)

In contrast to research, evaluation comprises the determination of criteria, setting standards on these criteria, data collection to inform criteria, and synthesis of the descriptive and factual information with the criteria to enable decision making about the object under evaluation (i.e., evaluand) within a set timeframe. While knowledge generation is part of many evaluations and usually a side effect, it is not the primary concern.

*Evaluation OF sustainability*

The determination of the merit, worth, and significance of efforts to continue a given evaluand (i.e., evaluation object) beyond the removal of initial program resources: What is the level of sustainability of your evaluand? How well is the evaluand sustained? Should it be sustained?

*Evaluation FOR sustainability*

The determination of the merit, worth, and significance in maintaining, replicating, and exporting a given evaluand’s positive (un)intended outcomes and impacts under specific consideration of global sustainability issues. How well does the evaluand
contribute to sustainable development efforts (human, social, economic, and environmental dimensions)?

**Checklist**

A tool that guides evaluation efforts, also known as, a framework for conducting evaluation. The SEC specifically can be compared to a heuristic that encourages critical thinking about sustainability evaluation.

**Guidance: How to use the SEC?**

The flow chart below is intended to provide guidance on how to use the checklist. You may also screen the summary table (after the flow chart) to determine whether specific aspects of the checklist might be of relevance to you.
SEC Flow Chart
Evaluation with an emphasis on sustainability requested

Don't use this checklist

Yes

Are you clear about the general considerations in evaluation?

No

Not sure

Check Section 1

Yes

Do you understand the object under evaluation in terms of sustainability?

No

Not sure

Check Section 2

Yes

Are you clear about the general logic of evaluation?

No

Not sure

Check Section 3

Yes

Are you clear what to consider to learn whether the evaluation object is important?

No

Not sure

Check Section 4

Yes

Are you clear about the properties that define good sustainability?

No

Not sure

Check Section 5

Yes

Are you clear about what to consider to learn whether continuation is worth the costs that accrue now and in the future?

No

Not sure

Check Section 6
SEC Quick Guide

Please consider this summary table to determine which aspects to further explore in the SEC.

<table>
<thead>
<tr>
<th>Part A – General considerations in evaluation</th>
<th>Clear</th>
<th>Needs consideration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1: Grounding the EVALUATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direction of the evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User of findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose(s) of the evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roles of the evaluation team</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Composition of the evaluation team</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timeframe under evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Key questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metaevaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissemination of findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 2: About the EVALUAND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The evaluation object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Components of the evaluation object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustainability dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local historical Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stakeholders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impactees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 3: General procedures for evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collect data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synthesize</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part B – Criteria</td>
<td>Not relevant</td>
<td>Relevant</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Section 6: Significance: Is the continuation of the evaluand important?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs for human sustainability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs for social sustainability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs for economic sustainability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs for environmental sustainability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scope and duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risks and vulnerability</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Section 4: Merit (Quality): What are the properties which define good sustainability without consideration of cost?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of evidence from monitoring and evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appreciation of knowledge, skills, abilities, competencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership competencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizational characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaboration/involvement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding the community and its environmental context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal orientation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive and negative impacts over time</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Section 5: Worth: Is the continuation of the evaluand or its outcomes worth the costs that accrue now and potentially will accrue in the future?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time at which costs/resources are accrued</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stakeholders and impactees, to whom monetary and nonmonetary costs accrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facets of cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific costs or resource use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource renewal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Notes:

- Sustainability can be evaluated as an element in its own right or as part of a more comprehensive evaluation. Linkages to process, outcome, and impact evaluations exist. Sustainability is inherent in:
  - Persistence of the institution
  - Persistence of program activities, services, interventions (this includes transferability to other contexts or replication of programming)
  - Persistence of resulting changes for individuals (humans), society (e.g., culture, institutions, etc.), economy, and the environment

- Not all components in the checklist are relevant in all cases or for each unique manifestation of sustainability. It is your responsibility to consider what is or is not relevant for a given situation. You may want to engage in discussions with stakeholders and impactees to determine which aspects are of special importance in your case.

- The SEC does not denote levels of sustainability performance, because these levels may vary for each sector, type of project, and region, nor does it prescribe the use of each checkpoint. Instead it suggests a holistic strategy to evaluation planning and design with an emphasis on sustainability evaluation, which often comprises a part of an evaluation, rather than an evaluation on its own.

- Efforts will be made to maximize the checklists user-friendliness for all individuals. Special efforts will be made to address the need of people with disabilities. Your feedback on how to best serve you as well as other groups is very welcome.

- The SEC will be posted online at www.sustainabilityeval.net

  Efforts will be made to interlink items to better reflect cross-linkages inherent in the concept sustainability as well as in the process of evaluation

- The SEC is a continuous work in progress and will be revised according to new insights and feedback. If you use the SEC, please share your experience with me:
  - What works?
  - What doesn’t work?
  - Do you have any cases or examples that could be shared with users of the SEC?
Part A – General considerations in evaluation

Note: The following checkpoints should be discussed among evaluation team members, the evaluation client, and key evaluation stakeholders to clarify information needs, resources, methodological decisions, required levels of detail, and evaluation management. They comprise general considerations in evaluation that have been specified for sustainability concerns. In addition, you may also want to consider other evaluation checklists available at:

http://www.wmich.edu/evalctr/checklists/checklistmenu.htm

Section 1: Grounding the EVALUATION

These general considerations are of major importance for planning the evaluation. Most are general concerns that apply to any evaluation.

Direction: Who asked for the evaluation?

○ Bottom-up*: Was the evaluation initiated on the ground at the local project level (e.g., is driven by middle range or grassroots actors, focusing on a specific project)?

○ Top-down*: Was the evaluation requested by the donor/funder of the program (i.e., the evaluation begins on the highest level)? Sustainability may only be one dimension to consider among others (e.g., by the DAC criteria*)

○ Peer-evaluation: Is the evaluation initiated and implemented by recipients/participants/users of the evaluation object.

○ Mixed directions

Users: Who are the intended users of findings from the evaluation?

○ Evaluation-funders (e.g., donors)

○ Decision-makers

○ Individuals who are engaged in program implementation (e.g., administrators, staff, volunteers)

○ Current and potential participants/ recipients of the program/services

○ Partners/collaborators

○ Others (who else should know learn about the findings to maximize transparency and use of the evaluation)
Purpose(s): Why is the evaluation being conducted?

- Improvement-oriented (formative*): Is it the intent of the evaluation to improve the sustainability of an evaluand and/or its outcomes and impacts? What works? What does not work?
- Decision-making (summative*): Is it the intent of the evaluation to inform decision making about the program? Does the program meet the needs of its intended users? How do costs compare with the benefits? Can outcomes be linked to the programming?
- Accountability*: Are funds being used for intended purposes? Has the program been implemented as designed?
- Knowledge generation (ascriptive*): Is it the intent of the evaluation to generate knowledge about: How sustainability is manifest within an evaluand? Why certain aspects of an evaluand are sustainable or not? What elements contribute to or hinder sustainability? What factors affect the continuation of an evaluand and/or its impacts on sustainable development?
- Development*: How does the program affect sustainable development at large? What can the program impact and what not? What can it control or not control?
- Monitoring*: Is the program going smoothly? Is funding stable? Is participation increasing or decreasing?
- Multiple purposes

Roles of the evaluator/evaluation team: What is the role of the evaluation team?

- Internal evaluation team (e.g., staff members, funders, donors, clients, other stakeholders, participants/recipient)
- External evaluation team (e.g., independent consultants or other external entity)
- Mixed (e.g., the evaluator as a critical, but external, friend collaborating with internal members; collaborative evaluation)

Composition of the evaluation team: Who will and will not participate in the evaluation?

- What competencies are needed? (e.g., evaluation-specific, research methodology, statistics, qualitative data analyst, content area expertise,
sector specific expertise, transdisciplinary expertise, administrative support, etc.)

○ Where will the evaluation take place? (e.g., potential language constraints, need for translator, cultural expertise, local guide and cultural expert)

**Timeframe of the evaluand:** In what lifecycle stage is the evaluand (see Figure below)?

○ Prospective/Ex-ante evaluation:
  - Conceptualization and/or development: Sustainability is being considered pro-actively

○ Prospective/Ex-ante evaluation AND/OR Retrospective/Ex-post evaluation
  - Growth and/or maturation: The level and breath of sustainability are considered
  - Reduction of initial funding resources: The stability of the evaluand and its outcomes and impacts are considered in terms of reduced resources or altered funding streams
  - Termination of initial funding resources: The stability of the evaluand as well as of the breadth and depth of outcomes and impacts are considered once initial resources have been terminated and funding streams adjusted to the new situation

○ Retrospective/Ex-post evaluation
  - After initial funding has ended: The stability of the evaluand as well as of the breadth and depth of outcomes and impacts are considered months, ideally years, after initial resources have been terminated and funding streams adjusted to the new situation (see termination of initial funding resources, the same considerations in a longer time frame)
Prospective Evaluation for Sustainability:
Emphasis on context, input, processes, outputs en route

Retrospective Evaluation of Sustainability
Emphasis on outcomes/impact beyond the immediate reach and life cycle

Key questions: What needs to be understood or learned about?

- Evaluation OF sustainability:
  - What is the level of sustainability of the evaluand at this time?
  - How well is the evaluand sustained?
  - Should sustainability be maintained?

- Evaluation FOR sustainability:
  - How well does the evaluand contribute to sustainable development in the long term?
  - How can sustainability of the evaluand be maximized?

Type of evaluation: How will you undertake your sustainability evaluation?

- Holistic: Do you need to have a general understanding about the sustainability of your evaluation object?
  - The breadth and depth of sustainability of the whole evaluand is considered without separating parts and/or dimensions
  - Usually expert evaluation
Analytic: Do you need to learn about the sustainability of specific components or dimensions of your evaluand? Or do you need to know whether the underlying theory works in terms of sustainability?

- Component evaluation: For example, an international development program may be implemented in different locations, have varying activities, and/or use differing policy instruments. Each element is evaluated separately before making judgments about the program as a whole.

- Dimensional evaluation: Sustainability can be treated as a dimension in its own and could be combined with other dimensions of importance in international development (e.g., relevance, cost-effectiveness, efficiency, efficacy, etc.)

- Theory-driven evaluation: Based on a program’s logic model, linkages between inputs, activities, immediate, intermediate, and long-term outcomes and impacts are examined to determine whether assumptions about the program are correct and if causality between program elements exists

- Mixed forms of the above

**MetaEvaluation**: Will the evaluation be evaluated? How will it be evaluated and by whom?

- Utility: Does the evaluation meet the information needs of intended users?
- Feasibility: Is the evaluation practical, politically viable, and cost effective?
- Propriety: Is the evaluation legally and morally sound?
- Accuracy: Is the evaluation technically sound?

- Note: Evaluation standards exist in several nations and organizations. These are often good points of references for considering the quality of an evaluation. This checklist could also serve as a point of reference on how well a given sustainability evaluation was conducted, specifically if the evaluation considered the critical element in sustainability evaluation.
**Dissemination of findings:** How are findings presented to users and other audiences? What presentation format(s) will facilitate learning for intended users?

- Technical report
- Briefing papers
- One page summaries/memos
- Presentations/workshops
- Other

**Evaluation management**

- Costs and resources of the evaluation
- Time available for the evaluation

---

**Section 2: About the EVALUAND**

This information is important for understanding the evaluand and its context. Checkpoints in this section are intended to clarify the nature of the thing under evaluation and its context. The key questions are: What, where, how, by whom?

**Types of evaluation objects (evaluand):** What is being evaluated?

- Policy*
- Proposal*
- Project*
- Program*/Project*
- Process*: Persistence of activities/services
- Product/results/outcome/impact*: persistence of changes
- Performance*
- Organization*/Institution*
- Other

**Components of evaluation objects:** What does the evaluation object consist of?

- Inputs* (monetary and nonmonetary resources, funding sources, in-kind contributions, technology, etc.)
- Activities/services/strategies
Outputs*/results (e.g., numbers who receive services/participated in activities)

Outcomes (e.g., changes as a result of participation in activities, of receiving services, of collaborating, etc.)

Potential (un)intended impacts (e.g., long-term intended and unintended effects on people locally, regionally, nationally, etc.)

**Sustainability dimensions of evaluation objects:**

- **Evaluation OF sustainability**
  - Project-oriented* sustainability: continuation of components (see above) in the interest of the funder or for local use
  - Purpose/use-oriented* sustainability: adaptation of an evaluation object or components thereof for local use
  - Systems-oriented* sustainability: integration of an evaluation object or components thereof to improve performance across a system (institutionalization, routinization)
  - Behavior-oriented* sustainability: capacity to effectively adapt to environmental changes

- **Evaluation FOR sustainability**
  - Human* sustainability: maintaining human capital such as health, education, knowledge, leadership, etc.
  - Social* sustainability (organizations and networks): maintaining social capital: cultural, language, shared rules, laws, etc.
  - Economic*(financial) sustainability: keeping capital intact
  - Natural (environmental*) sustainability: Protecting natural capitals (e.g., water, land, air, minerals, etc.)

**Local historical context:** What is the local historical evolutionary context?

Have there been significant changes in the past?

- Human: Culture, beliefs, perceptions
- Social: Collaboration and partnerships
- Economic: Nature of economy and investment patterns, funding agencies, community assets
- Environmental: Local/regional challenges
- Politics: Political support and risks
o Administrative: binding acts, decision makers commitment
o Technological innovation
o Resource availability: what natural, physical, human, monetary and nonmonetary (e.g., time) resources are available or not

<table>
<thead>
<tr>
<th>Context</th>
<th>Past</th>
<th>Current</th>
<th>Projected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultural</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technological</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stakeholders*: Who can affect the evaluand and its sustainability?

o Internal stakeholders:
  - Individuals involved in implementing the project/program (e.g., staff, volunteers, partners)
  - Funders, governments, NGO’s, etc.
  - Users/participants/recipients82
  - Others

o External stakeholders:
  - Supporters
  - Politicians
  - Dissidents/Protestors/Oppressors
  - Alternative stakeholders: those who could affect the evaluand and its sustainability either positively or negatively

Impactees* of the evaluand: Who is affected by the evaluand?

o Internal impactees

---
82 Commonly referred to as “beneficiaries.” The terms “recipient(s), participant(s), and user(s) are preferred here to assure that benefits are not falsely attributed.
- Individuals involved in implementing the project/program (e.g., staff, volunteers, partners)
- Funders, governments, NGO’s, etc.
- Users/participants/recipient
- Others
  - External impactees
    - Family member, friends, business partners, colleagues, community members at large, attentive audiences, consumers
    - Alternative impactees: those who could have been impacted or protected from impact

*Reach* of the evaluand: How far do impacts potentially reach?

- **Space:**
  - Local impacts on people
  - Sub-national impacts (i.e., multi-province, county, state, etc.)
  - National impacts on people
  - Multinational regional impacts on people
  - International (global) impacts on people

- **Time**
  - In the past
  - In the present (now)
  - In the immediate future (1-10 years)
  - In the intermediate future (over a life time)
  - In the long-term (future generations)

**Section 3: General procedures for evaluation**

These procedures lay out the general logic of evaluation. It is up to the user to determine the best models and approaches to answer the specific questions in her context (i.e., choose the working logic). This also relates to the types of data collected and used within the evaluation. Ideally, an evaluation would employ mixed methods though the questions should determine the most appropriate methods in a given evaluation setting. This section builds on the previously clarified information about the evaluation and the evaluand.
**Identify criteria**: On what components or dimensions must your evaluand do well to be considered of good quality, value, and importance?

- Specific sustainability related criteria are listed in Part B
- Additional criteria of relevance to your evaluation object can be identified via:
  - Needs assessment: What are the most critical needs in the community that can be addressed by the evaluand?
  - Assets and opportunity assessment: What assets and opportunities are in the community to meet the needs?
  - Risk/ vulnerability assessment: What risks and vulnerabilities may prevent sustainability?
  - If no recent and/or valid assessments are available, conduct one to determine human, social, economic, and environmental needs and risks that the evaluand ought to address

**Identify values and set standards (e.g., targets)**: What constitutes good/bad, worthless/worthwhile, relevant/not relevant? Are there grey areas? What is acceptable and what not? How will you know? How do we know what good means?

- What constitutes good/bad, worthless/worthwhile, relevant/not relevant? Which shades of gray exist? What is acceptable and what not? How will you know? Strategies to find out include:
  - Exploring organizational values: Are there any predetermined values and targets specified by the organization(s) responsible for the evaluand?
  - Considering the assets, needs, risk, and vulnerability assessment(s): what are the most severe needs, risks, and vulnerabilities and what community assets are available to relieve them?
  - Knowledge of legal and policy documents
  - Ethical standards
  - Human rights
  - Other

- Are some criteria more important than others? (weighting the relative importance of the criteria) Strategies to find out if this is the case include:
- Having stakeholders and/or impactees vote
- Using the knowledge of selected stakeholders or experts
- Employing evidence from the literature
- Using evidence from the needs, assets, vulnerability, and assets assessments
- Using logic modeling and evidence of causal linkages

- Should criteria be graded* or ranked*?
  - Grading (rating): Assigning the evaluand or its components or dimensions to an ordered set of categories, with the order corresponding to a metric of merit, worth, and/or significance
  - Ranking: Placing the evaluand or its components or dimensions in an order of merit worth, and/or significance on the basis of their relative performance on a measurement or observation.

- What constitutes minimum acceptable standards (bars*) and other performance standards*?
  - Given the resources invested in the evaluation object, is there a minimum level of sustainability that has to be achieved on a given component, dimension, and/or overall?
  - Rubrics* can help explain how performance on the criteria will be rated: In your case, what does it mean to perform inadequately, adequately, or exceptionally (you may want to choose a different type of scale)
  - Use the strategies above to inform minimum standards and rubrics

**Collect data (measure/observe criteria) and compare with the standards:**
How well did the evaluand perform?

- What do you need to know to make decisions about how well the evaluand performs on a given criterion?
- Are there any indicator* sets that inform the criteria of interest? You may want to consult indicator frameworks that have been developed by the sustainable development community (e.g.,
- Who can provide the information?
  - People (see your list of stakeholders and/or impactees; i.e., consider those whose opinions may not be reflected in available written documents)
- Organizations
- Documents, the literature, previous evaluations, data from monitoring
- Others
- Note: Ask yourself for rationales for including and excluding specific information sources. Different stakeholders and impactees may have very differing perceptions about what is good, worthwhile, and important and bring differing perspectives
  - Determine how to collect and analyze the needed data
    - By what means will you get the needed data: document and literature reviews, observations, tests, questionnaires, interviews, focus groups, site visits, and/or other methods of data collection
    - Are data collection instruments available?
      - If not, develop instruments with input from key informants (consider your list of stakeholders and impactees)
      - Collect data on the quality of the instruments (reliability and validity) – remember that the product of the evaluation can only be as good as the technical rigor
    - Is it possible to collect data from varying sources to allow for triangulation of information and perspectives
    - Assure that data analytic strategies are adequate (reliable, valid, credible)
  - Consider time and space dimensions of sustainability (cross reference)

**Synthesis**: Integrate data with the standards on criteria into a judgment of merit, worth, and/or significance
  - Integrate the data (ratings or grading on criteria) with the standards
  - Depending on whether you are doing a holistic or analytic evaluation, determine how the relative or absolute merit, worth, and/or significance will be determined
    - Grading and ranking provide unidimensional conclusions about components or dimensions
• Profiling* provides multidimensional conclusions, usually depicting grades and comparative performance on components or criteria
  o Ensure that evaluative conclusions (claims) are legitimate
  o Identify strengths and weaknesses of the evaluative conclusions

Part B – Criteria of sustainability for sustainability

The following sections list potential criteria of significance, merit, and worth in sustainability evaluation. In contrast to indicators, criteria of sustainability are those properties of an evaluation object that are part of good sustainability in a given context.

Section 4: Criteria of significance (importance)

Is the continuation of the evaluand important for sustainable development?

These criteria are specifically concerned with the relevancy of the investment. What is found to be important may vary by stakeholder and impactee groups and can differ by evaluand.

Needs for human sustainability*:
  o Nutrition, shelter, clothes
  o Education, health, means of transport and communication, safety
  o Belongingness, creativity, identity, autonomy, spirituality
  o Togetherness, participation
  o Self-fulfillment
  o Realization of potential
  o Other

Need for social sustainability*:
  o Social norms, community cohesion for mutual benefit
  o Connectedness between groups of people
  o Cultural plurality
  o Solidarity
o Tolerance, respect, compassion, patience, and honesty
o Discipline
o Commonly shared rules, laws, and information
o Equity across gender, age, religions
o Human rights
o Peace
o Participation in decision-making about planned interventions that affect people's lives
o Justice
o Accountability
o Self-reliance/dependency: specifically mobilization of communities, local ownership in decision making, commitment of local resources
o Politics

Needs for economic sustainability:
o Economic benefits to impactees and stakeholders
o Reduced need for external assistance
o Allocation of financial resources
o Efficiency
o Scale of consumption
o Preventive anticipation
o Cost-effectiveness under consideration of unduly costs
o Paying for past ecological debt
o Optimizing productivity
o Use of human, natural and financial capital

Needs for environmental sustainability:
o Water, land, air, minerals, eco-system services
o Environmental soundness of the intervention, its intended and unintended outcomes and impacts
o Waste emissions within the assimilative capability of the environment without damaging it
o Ecological balance and biodiversity
o Balance in consumption/recycling of resources
o Disaster risk reduction
o Irreversible loss of species biodiversity, habitat, ecosystem
Scope and duration:
- Continuation of activities, service provision, or outputs
- Replication, transfer, or export of the evaluand
- Consider numbers and types of activities, services, and outputs as well as number of recipients/participants
- Duration
- Adaptation (at what point is that what we intend to sustain changed to the degree to which we cannot call it the evaluand anymore?)

Risks*/Vulnerability: Unawareness of risks may thwart sustainability. It is critical to be aware of potential risks to the sustainability of the evaluand that threaten potential for long-term sustainability. What strategies are in place to thwart potential risks to successful continuation of relevant and successful components and dimensions of the evaluand?
- Flexibility to changes in the environment
- Cultural compatibility of activities
- Risk to human sustainability: e.g., overpopulation and human development
- Risks to social sustainability: e.g., violence and social breakdown
- Risk to economic sustainability: e.g., crisis and shocks, balance of payments
- Risk to environmental sustainability: e.g., climate change, natural disasters (earthquakes, tsunamis, etc.), over-consumption, waste, etc.
- Risk to participants and program staff if the evaluand is or is not sustained

<table>
<thead>
<tr>
<th>Detriment-Benefit Matrix (adapted from Davidson, 2005a)</th>
<th>How detrimental would it be overall if the evaluand did very poorly on this dimension of sustainability?</th>
</tr>
</thead>
<tbody>
<tr>
<td>How beneficial would it be overall if the evaluand did very well on this dimension of sustainability?</td>
<td>Not noticeably detrimental</td>
</tr>
<tr>
<td>Somewhat beneficial</td>
<td>Desirable*</td>
</tr>
<tr>
<td>Very beneficial</td>
<td>Desirable*</td>
</tr>
<tr>
<td>Extremely beneficial</td>
<td>Important**</td>
</tr>
</tbody>
</table>
Section 5: Criteria of merit (quality)

What are the properties that define good sustainability without consideration of cost?

Process-oriented criteria: These criteria are especially useful in determining if the evaluand has the capacity for addressing sustainability needs (i.e., prospective considerations), but also to determine which of the evaluand’s elements (e.g., activities) persist after initial resources for the evaluand have been removed (i.e., retrospective considerations) or supplemented with other resources.

Use of evidence from research, monitoring, and evaluation
- Not everything should be sustained
- Not everything was planned to be sustained
- Factors, activities or outcomes to be sustained were found to be adequate to allow for the continuation of the evaluand
- Findings from monitoring and evaluation are used proactively for continuous improvement of the evaluation object
- There is evidence that factors that have shown to be detrimental or insufficient have been discontinued or have been improved

Appreciation of knowledge, skills, abilities, competencies
- Traditional/classical knowledge
- Intra- and intergenerational knowledge
- Knowledge management
- Accessibility to knowledge: Awareness of international values and policies in the international community (Millennium Development Goals, the United Nation’s Development Assistance Framework, and Agenda 21)

Leadership competencies
- Championing: Capacity to promote sustainability, while preventing or mitigating negative impacts.
- Strategic plans for sustainability are frequently revised to adjust for changes in context. Activities and goals are aligned with sustainability needs of consumers and impactees (see criteria of significance)
- Strong political commitment and external support are obtained
o Consensus and long-term vision about objectives
o Shared strategic and pragmatic vision
o Realistic and flexible targets
o Succession planning
o Commitment to sustainable development
o Balance between bureaucratic efficiency and democratic involvement (i.e., effective participation)

Organizational characteristics (replicability of sustainable actions and/or impacts)

o Relevant knowledge, skills, and abilities of those involved
o Diversified funding streams
o Diversified activities
o Participation, involvement, and integration at all levels
o Equity
o Institutionalization efforts are in place
o Continuous monitoring of progress toward sustainability
o Legal basis of the organization

Infrastructure: Reach, condition, and match of infrastructure in relation to program/project goals.

o Adequacy of technology (e.g., communications, mobility)
o Accessibility of people to be reached (e.g., roads, vehicles)
o Stable electricity, if needed
o Adequate waste treatment
o Adequate facilities for activities, program support, etc.

Collaboration/Involvement

o Inclusion of relevant stakeholders and impactees
o Clear communication/transparency (e.g., sharing vision, sharing findings from evaluations)
o Linkages to other organizations/partners
o Communication patterns among participants on the local, national, and international level of the evaluand and respective knowledge transfer (systemic support mechanisms)
Collective responsibility and accountability

Understanding the community and its environmental context: Is there a conducive environment for sustainability?

- Respect (sensitivity) for the community’s tolerance for change (e.g., cultural relevance, respect of indigenous knowledge and practice; enablers and inhibitors to sustainability)
- Acceptability across impactees (i.e., consider impacts on men/women, young/old, healthy/sick)
- Politics and power relationships
- Appropriateness of policy frameworks (e.g., people-centered, rights-based, community-driven development model)
- Alignment of intervention with local, national and/or international policies and priorities
- Alignment of intervention with legal requirements: What federal, national, or state laws and regulations are relevant across program boundaries? How could these laws/regulations affect sustainability?
- Recognition and preservation of diversity
- Community linkages (social capital: willingness to share knowledge and information, help in handling everyday matters, and reinforcing social networks, solidarity)
- Appropriateness of technology (e.g., simplicity, affordability, adequacy, mobility, etc.)

Responsiveness

- Awareness of current and emergent needs
- Ability in addressing emergent needs within the realm of the organization’s mission and priorities
- Ability to adjust to changing contexts
- Ability to adjust for unanticipated negative impacts and side effects (e.g., environmental degradation)
- Substitution of resources in contrast to usage of nonrenewable resources
- Continuous adaptation of intervention to maximize benefits and minimize harm
- Concern of potential harms of an intervention to future generations (intergenerational equity; inclusion of children and youth specifically)
- Integrated renewal mechanisms

**Outcome-oriented criteria:** These criteria might be especially useful in determining if the evaluand has the capacity for sustainability (i.e., prospective considerations) or which outcomes have been sustained to date (i.e., retrospective considerations). Remember that not everything should be sustained.

**Goal orientation (key issues in sustainable development):**
- Was the evaluand intended to be sustained?
- Was the evaluand implemented as designed?
- Consideration of the whole system and its parts: linkages between interventions and outcomes
- Consideration of human, social, economic, and environmental subsystems (holistic science and appropriate technology)
  - Human subsystems: consider for example developing full human potential; cultural, moral and spiritual sensitivity; self-determination; population growth
  - Social subsystems: consider for example gender sensitivity; social justice; tribal ecological knowledge
  - Economic subsystems: consider for example institutional viability; viable, sound and broad-based economic
  - Environmental subsystem: consider ecological soundness

**Positive and negative impacts of the evaluand over time:**
- Human dimension: consider for example health, education, poverty reduction; availability and quality of food
- Social dimension: consider for example politics, local partnerships, gender, age, equity; ethics; cultural beliefs, language, values; indigenous rights; community cohesion, stability, character, services, and social institutions; politics; impacts on **tribal ecological knowledge**
- Economic dimension: consider for example access to and control over resources; infrastructure, institutions, tourism
- Environmental dimension: consider for example aesthetics (landscape analysis); archaeology and heritage; quality of air and water; level of
exposure to risk, hazard, noise, dust; the local ecological condition on which life depends

- Cross-dimensional impacts: consider for example intergenerational and intra-generational equity (gender, age, race/ethnicity, etc.); empowerment via social mobilization, direct action, power or protest-based confrontation, economic and social production approaches, civic engagement, raising consciousness, building capacity by providing knowledge, skills, and positive experience/success; usefulness of outcomes to community; the precautionary principle

Section 6: Criteria of worth (value)

*Is the continuation of the evaluand or its outcomes worth the costs that accrue now or potentially will accrue in the future?*

Available cost evaluation checklists provide guidelines for identifying and analyzing costs and benefits that can be useful in many evaluations. In evaluation of sustainability, however, the consideration of cost must not only reflect on benefits and costs but also entail current and future generations. Benefits from an intervention should outweigh costs to human and the environment.

**Time at which costs/resources are accrued**

- During the general program life cycle: What monies and resources are, were, and could have been used during the program life cycle?
  - This information should be available from previous evaluations, monitoring activities, or program documentation based on which the sustainability evaluation has been deemed worth doing
  - Consider adequacy of financial, human, material, and other resources. This will impact the scale and scope of the evaluand

- After the termination of startup funding: what monies and resources are, were, and could have used to continue elements of the evaluand? These monies and resources show the capability of the evaluand to continue

- Time of future generations (e.g., 20 years later): what monies and resources are needed to continue important elements of the evaluand in
the long run? These costs are those resources and capacities required for maintaining the evaluand and outcomes thereof over time

**Stakeholders and impactees, to whom monetary and nonmonetary costs accrue** (check your list of consumers and impactees developed earlier to make sure that you do not forget anyone or any group of people or organization that is of importance here)

- Costs that accrue for those people involved in the programming. These costs include payments/salaries, time, resources, but also personal costs (e.g., stress, time away from the family, etc.)
- Costs that accrue for those people that are impacted by the programming. These include intra-generational and intergenerational impactees and upstream stakeholders as well as alternative impactees.

**Facets of cost (generally, monetary and nonmonetary are distinguished here):**

- Actual costs to humans (monetary and nonmonetary costs that accrue to individuals)
- Actual costs to society (monetary and nonmonetary costs that accrue to groups of people, organizations, communities, etc)
- Opportunity costs*: the cost of not considering alternatives at each of the previously stated levels

**Specific costs or resource use to consider include:**

- Human resources
- Renewable and nonrenewable resources
- Tools and technologies used
- Infrastructure
- Recycling, waste management, and conservation
- The benefits from the evaluand are equal to or larger than the costs accrued

**Resource renewal**

- Diversification of funding
- Maximization of assets
| A | Activity/activities | Actions that are assumed by an evaluation object to achieve goals (see Frechtling, 2007) |
| A | Analytic evaluation | Components, dimensions, and/or the underlying theory of the evaluand are considered separately prior to synthesizing these ‘subevaluations’ into conclusions about the whole evaluand (see Davidson, 2005a) |
| A | Apportioning | Allocation or distribution: Dividing a given, often finite, quantity of valued resources between competing demands |
| A | Ascriptive evaluation | Evaluation for the purpose of knowledge gain (for the sake of it) |
| A | Audience | Users of the evaluation; those who should receive reports, presentations, workshops, debriefings, etc |
| B | Bar | Minimum acceptable standard; an evaluative operation where minimum levels of performance are set, or required, on specific dimensions or components, performance below which cannot be compensated for by better performance on other dimensions. Failure to ‘clear’ a bar means ‘failure’ of the evaluand |
| B | Behavior-oriented sustainability | The target group or project holder has problem-solving capacities to adequately and flexibly adapt to changing environmental conditions (after Stockmann in Caspari, 2004, pp. 67-68) |
| B | Benefit | Positive outcome or impact |
| B | Bottom-up evaluation | The evaluation is initiated on the grassroots level |
| C | Checklist | A tool to plan and design evaluation |
| C | Component evaluation | Each part of the evaluation object (e.g., inputs, activities, outputs, outcomes) is evaluated separately. The resulting subevaluations are then integrated into overall conclusions about the evaluand (see Davidson, 2005a). |
| C | Criteria | Properties that are part of the concept of a “good X;” they are definitionally connected with the evaluand (see Scriven, 2007) |
| D | DAC-Criteria | Principles for evaluations of development interventions (OECD, 2007) |
### Dimensional Evaluation

Merit, worth, and significance are considered for facets that permeate the whole evaluand; in sustainability evaluation, these facets include human, social, economic, and environmental dimensions.

<table>
<thead>
<tr>
<th><strong>E</strong> Economic dimension of sustainability</th>
<th>Financial stability: keeping capital intact; concerns economic needs, infrastructure, distribution of wealth, control over resources, overconsumption, etc. (cf., Goodland, 2002)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Environmental dimension of sustainability</strong></td>
<td>Protecting natural capitals (e.g., water, land, air, minerals, etc.); Concerns ecological needs, pollution, climate change, waste management, green energy, etc. (cf., Goodland, 2002)</td>
</tr>
<tr>
<td><strong>Evaluand</strong></td>
<td>Something that is being evaluated, object under evaluation (e.g., products, policies, programs)</td>
</tr>
<tr>
<td><strong>Evaluation</strong></td>
<td>The systematic process of determining the merit (quality), worth (value), and/or significance (importance) of evaluands (e.g., programs, policies, and products) or evaluatees (e.g., personnel), or the product thereof. Professional evaluation involves the use of systematic investigation to collect and synthesize factual information (what so?) to render evaluative conclusions (so what?) about an evaluand’s goodness, value, and importance.</td>
</tr>
<tr>
<td><strong>Evolutionary sustainability</strong></td>
<td>A model assuming that sustainability relies on an interplay between forecasting, retaining, and adapting organizational processes (cf., Kraft &amp; O’Neill, 2007)</td>
</tr>
<tr>
<td><strong>External evaluation</strong></td>
<td>Evaluation conducted from outside an organization or program; the evaluator is not on the pay role of the organization that designed or implemented the program (see Davidson, 2005a)</td>
</tr>
<tr>
<td><strong>F</strong> Formative evaluation</td>
<td>Improvement-oriented determination of merit, worth, and/or significance with the intent to inform decision making about the state of an evaluand’s/evaluatee’s (e.g., program, policy, personnel) components or dimensions. This type of evaluation supports decision making about which program components or staff member competencies require improvement.</td>
</tr>
<tr>
<td><strong>G</strong> Global bar</td>
<td>Involves setting or requiring minimum levels of combined or aggregated performance across all dimensions or components, normally using numeric</td>
</tr>
</tbody>
</table>

---

**Note:** The table above is an extract from a larger document, focusing on the concepts of dimensional evaluation, economic and environmental dimensions of sustainability, evaluand, evolution, and external evaluation. Each cell provides a brief explanation or definition related to these concepts.
<table>
<thead>
<tr>
<th><strong>Index</strong></th>
<th><strong>Definition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal-free evaluation</td>
<td>The determination of merit, worth, and/or significance without explicit consideration of a program’s stated goals or objectives. Goal-free evaluation considers what an evaluan is doing instead of what it intended to do. Needs assessments are central elements in goal-free evaluation.</td>
</tr>
<tr>
<td>Grading</td>
<td>Assigning evaluan to an ordered set of categories, with the order corresponding to a metric of merit</td>
</tr>
<tr>
<td>H Holistic bar</td>
<td>Involves a visual inspection (i.e., non-numeric) of performance across all dimensions or components, where performance across all must meet a minimum in order to ‘pass’ (see Coryn, 2006).</td>
</tr>
<tr>
<td>Holistic evaluation</td>
<td>The whole evaluan is considered without separating parts</td>
</tr>
<tr>
<td>Human dimension of sustainability</td>
<td>Maintaining/improving human capital such as health, education, knowledge, leadership, etc.; Concerns basic human needs such as food, shelter, health, etc. (cf., Goodland, 2002)</td>
</tr>
<tr>
<td>Impact</td>
<td>Intended, unintended, anticipated, and unanticipated effects on targeted and non-targeted populations; usually referring to long-term effects and outcomes (see Davidson, 2005a; Frechtling, 2007)</td>
</tr>
<tr>
<td>Impactee</td>
<td>Everyone who experiences change due to the evaluan, including individuals who are directly affected by an intervention (i.e., downstream direct impactees) and individuals and organizations that are NOT involved in the evaluan and are NOT direct recipients of the evaluan, but are still impacted by the potential range of outcomes of the evaluan recipients (i.e., downstream indirect impactees), those directly involved in the program implementation (e.g., staff; i.e., mid-stream consumers), and funders, political supporters, etc. (i.e., upstream impactees), (see Scriven, 2006)</td>
</tr>
<tr>
<td>Indicators</td>
<td>Factors, variables, or observations that provide evidence for the performance on a given criterion. Sustainability indicators must be specific (relate to the criterion), measurable (or observable), usable (practical), sensitive (must readily change as circumstances change), available (data must be collectable); and cost-effective (see Bell and Morse, 2003)</td>
</tr>
<tr>
<td>Input</td>
<td>Material, non-material, monetary, and non-monetary resources of an evaluation object</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Internal evaluation</td>
<td>Evaluation conducted from within an organization or program; includes self-evaluation and evaluation by peer from different programs or units within the same organization (see Caspari, 2004, p. 32); the evaluator is on the pay role of the organization that designed or implemented the program (see Davidson, 2005a)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M</th>
<th>Merit</th>
<th>Intrinsic quality; quality without consideration of cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metaevaluation</td>
<td>Formative or summative evaluation of evaluation processes and products. Standards against which evaluations can be assessed include, for example, the Joint Committee Standards for Program Evaluation, the American Evaluation Associations Guiding Principles for Program Evaluators, and the U.S. Government Accountability Office’s Government Auditing Standards (Yellow Book).</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mnemonic device</th>
<th>Memory aid</th>
</tr>
</thead>
</table>

| N | Needs assessment | A systematic approach for determining states of existence or levels of performance of people, programs, or organizations. The purpose is to set priorities, allocate resources, and/or determine evaluative criteria. In contrast to wants or ideals, needs are essential for people, programs, or organizations to exist and perform reasonably in a given context. When conducting needs assessments, it is important to consider (un)met and (un)conscious needs as well as performance and treatment needs. |

<table>
<thead>
<tr>
<th>O</th>
<th>Omission</th>
<th>Oversight, exclusion (Scriven, 1991; 2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunity cost</td>
<td>Activities or services that could have been implemented if resources had been allocated differently; forgone alternatives (Davidson, 2005a; Mathison, 2005)</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td>Usually intended, but also unintended change occurring as a consequence of the evaluand’s activities, progress toward goals (Frechtling, 2007; Davidson, 2005a)</td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td>Tangible, immediate results that are evidence for the implementation of an activity or service (Mathison, 2005; Frechtling, 2007)</td>
<td></td>
</tr>
</tbody>
</table>

<p>| P | Performance | A factor determining overall productivity or functionality of an evaluand or evaluee |</p>
<table>
<thead>
<tr>
<th><strong>Performance standards</strong></th>
<th>Performance standards are specific values applied to the general criteria. They clarify what comprises different degrees of 'good,' 'valuable,' and/or 'important.'</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Policy</strong></td>
<td>Written plans that are informed by evidence, change focused, inclusive, strategic, causal, realistic, flexible, and outcome oriented (see Owen, 2006, p. 26) with the intend to guide decision making and action in specified contexts (see Davidson, 2005, p. 244). Policies can exist on local, national, and international levels.</td>
</tr>
<tr>
<td><strong>Process</strong></td>
<td>That what is being implemented within the realm of a program, including consideration of inputs, activities, services, and outputs</td>
</tr>
<tr>
<td><strong>Product</strong></td>
<td>A concrete result of a performance, task, production, or other process including outputs, outcomes, and impacts</td>
</tr>
<tr>
<td><strong>Profiling</strong></td>
<td>To graphically exhibiting grades, not scores, on the relevant dimensions of merit, worth, and/or significance</td>
</tr>
<tr>
<td><strong>Program</strong></td>
<td>A set of planned activities or services intended to address a need or other goals for a specified target group (see Davidson, 2005a; Owen, 2006)</td>
</tr>
<tr>
<td><strong>Project</strong></td>
<td>See program; usually a small scale program that is more refined in terms of time and scope.</td>
</tr>
<tr>
<td><strong>Project-oriented sustainability</strong></td>
<td>The target group or project holder continues the evaluand in its own interest for its own purposes over time (after Stockmann in Caspari, 2004, p. 67)</td>
</tr>
<tr>
<td><strong>Proposal</strong></td>
<td>A written plan/offer that specifies a program or an evaluation, associated prices, terms and conditions, products, goals, etc.</td>
</tr>
<tr>
<td><strong>Purpose/use-oriented sustainability</strong></td>
<td>Other groups or project holders adapt the evaluand for their interests, purposes, and uses (after Stockmann in Caspari, 2004, p. 67)</td>
</tr>
<tr>
<td><strong>Ranking</strong></td>
<td>An operation used to place evaluands or valuees in an order of merit (worth or significance) on the basis of their relative performance on a measurement or observation.</td>
</tr>
<tr>
<td><strong>Reach</strong></td>
<td>The breadth of impacts resulting from an evaluation object geographically and over time</td>
</tr>
<tr>
<td><strong>Risk</strong></td>
<td>Factors that affect or are likely to affect success of an intervention; also negative consequences of an intervention to human life, health, property, or the environment</td>
</tr>
<tr>
<td>Rubric</td>
<td>Description of the meaning of a level of performance (e.g., inadequately, adequately, or exceptionally; scales can vary)</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>S: Scaling</td>
<td>Refers to replicating, expanding, or increasing program efforts to extend positive outcomes and impacts</td>
</tr>
<tr>
<td>Scoring</td>
<td>Involves assigning numeric quantities, usually in terms of performance, on which to represent merit</td>
</tr>
<tr>
<td>Significance</td>
<td>Importance, relevance</td>
</tr>
<tr>
<td>Social dimension of sustainability</td>
<td>Maintaining/improving social capital: cultural, language, shared rules, laws, etc.; concerns social needs, ways of organization, governance, and human interaction, etc. (cf., Goodland, 2002)</td>
</tr>
<tr>
<td>Stakeholder</td>
<td>Those with a stake or some sort of investment in the evaluation object (see Davidson, 2005a; Frechtling, 2007)</td>
</tr>
<tr>
<td>Summative evaluation</td>
<td>Accountability-oriented evaluation that seeks to determine the merit, worth, and/or significance of an evaluand in order to inform decision making about the evaluand. This type of evaluation aids decision making about whether to continue or terminate a program, or hire or fire a staff member.</td>
</tr>
<tr>
<td>Sustainability</td>
<td>The capacity to exist (e.g., projects, programs, mankind) or continue (e.g., human, social, economic, and/or environmental benefits).</td>
</tr>
<tr>
<td>Sustainable development</td>
<td>&quot;Development that meets the needs of the present without compromising the ability of future generations to meet their own needs&quot; (WCED, 1987, p.43).</td>
</tr>
<tr>
<td>Synthesis</td>
<td>The process of integrating a set of ratings or performances on several dimensions, components, or criteria into an evaluative conclusion</td>
</tr>
<tr>
<td>Systems-oriented sustainability</td>
<td>The evaluand is being implemented system-wide to enhance performance across the system (e.g., educational or health systems); (after Stockmann in Caspari, 2004, pp. 67-77)</td>
</tr>
<tr>
<td>T: Theory of an evaluand</td>
<td>Generally refers to the logic or guiding framework for a program; a set of assumptions about how a program works; also referred to as logframe, logic model, theory of change (see Frechtling, 2007)</td>
</tr>
<tr>
<td>Top-down evaluation</td>
<td>An elite group (e.g., the funder) is requesting the evaluation.</td>
</tr>
<tr>
<td>V</td>
<td>Value(s)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>W</td>
<td>Weighting</td>
</tr>
<tr>
<td></td>
<td>Worth</td>
</tr>
</tbody>
</table>