LOS ANDES UNIVERSITY

Developing a Systemic Program Evaluation Methodology: A Critical Systems Perspective

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS

For the degree

DOCTOR IN PHILOSOPHY

Engineering

By

Maria Alejandra Torres Cuello

BOGOTA, COLOMBIA

May 2018

© 2018

Maria Alejandra Torres Cuello

ALL RIGHT RESERVED

Developing a Systemic Program Evaluation Methodology: A Critical Systems Perspective

Abstract

Although the benefits of introducing systems thinking concepts at a theoretical level in the program evaluation field have been recognized, attempts in doing so have been very limited. For this reason, the following research seeks to address this issue by introducing systems thinking concepts in the methodological design of a program evaluation methodology. In doing so, a review of the most important systems thinking concepts as well as of the most prominent program evaluation models that have been developed so far is conducted. This is followed by choosing the systems concepts and the program evaluation methodology to be addressed. As a result, a reframed evaluation methodology is proposed, and issues related to power and judgmental biases in evaluation are explored from a perspective different from the one traditionally used to tackle these issues in the program evaluation field. This methodological development is later on put into practice by conducting the evaluation of the Hermes Program, a peace building program developed by the Bogota Chamber of Commerce in Bogota, Colombia. The description of the evaluation design, implementation, results, and recommendations is presented in this document.

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to Prof. Luis Pinzon and Prof. Gerald Midgley for the continuous support of my Ph.D. study and related research, for their patience, motivation, and immense knowledge. Their guidance and support, as well as that of their families, helped me in all the time of research and writing of this thesis.

I would also like to thank the rest of my thesis committee: Prof. Maria Elsa Correal, and Prof. Jose Rodrigo Cordoba for their insightful comments, encouragement and help throughout my research.

I thank Los Andes University and its administrative staff for giving me the opportunity of presenting and sharing my work in places and people that I would have never imagined.

My sincere thanks also go to Patricia Ricco, the director of the Hermes Program, and Claudia Marcela Alvarado who gave me the opportunity of working with the Hermes Program and provided the time and spaces for doing so. Without their support it would have not been possible to conduct this research. I would also like to thank the parents, students, teachers, administrative staff, and consultants of the Hermes Program who were committed with the development of different activities of this research. Without their time and commitment none of this would have been possible.

I thank every professor and student at the University of Hull who was involved with this research in one way or another. In particular Amanda Gregory and Maya Vashkova for their support, guidance, and for giving me a new perspective on my research.

Last but not least, I would like to thank God for always being my guiding light, my parents and closest friends for supporting, guiding, and encouraging me in the ups and downs not only in writing this dissertation but in my life in general as well.

TABLE OF CONTENT

Chapter 1 – Introduction	1
1.1 General Aim of the Thesis	1
1.2 Specific Aims of the Thesis	1
1.3 Relevance.....	3
1.3.1 Theoretical Relevance.....	3
1.3.2 Practical Relevance	7
1.4 Structure of the Thesis	8
Chapter 2 – Systems Thinking.....	11
2. 1 Precedent to Systems Thinking.....	11
2.1.1 The emergence of the system concept	11
2.1.2 General Systems Theory	12
2.2 The First Wave of Systems Thinking	14
2.2.1 Development of the first wave	15
2.2.2 Criticism of the First Wave of Systems Thinking.....	17
2.3 Second Wave of Systems Thinking	20
2.3.1 Developments of the Second Wave of Systems Thinking	21
2.3.2 Criticism of the Second Wave of Systems Thinking	22
2.4 Third Wave of Systems Thinking	24
2.4.1 The first strand of the third wave of systems thinking.....	25
2.4.1.1 System of Systems methodologies.....	28
2.4.1.2 Total Systems Intervention	30
2.4.2 Second strand of the third wave of systems thinking.....	35
2.4.2.1 Critical Systems Heuristics	35
2.4.2.2 Boundary critique, boundary judgments and marginalization	44
2.4.3 The third strand of the third wave of systems thinking.....	49
Chapter 3 – Evaluation Theory	52
3.1 The concept of evaluation	52
3.2 Evaluation Mapping.....	54
3.2.1 Mapping Methods	54
3.2.1.1 Evaluation Theory tree.....	55
3.2.1.2 The Four Generations of Evaluation.....	59

3.2.1.3 Justification for these classification forms	62
3.3 The proposed evaluation mapping for this research	64
3.4 Evaluation Waves	66
3.4.1 First Evaluation Wave – Setting the ground for the formalization of the evaluation field	66
3.4.2 Second Evaluation Wave	69
3.4.2.1 Campbell & Stanley’s Quasi Experimental and Experimental Designs	69
3.4.2.2 Boruch and the Randomised Field Experiment	70
3.4.2.3 Lee Cronbach and the UTOS Development.....	72
3.4.2.4 Levin, McEwan and Cost-Effectiveness analysis	73
3.4.3 Third Evaluation Wave	75
3.4.3.1 Rossi, Chen & Weiss and the development of Theory Driven and Theory Based Evaluation .	75
3.4.3.2 Daniel Stufflebeam and the CIPP Model	77
3.4.3.3 Patton’s Utilization Focused Evaluation.....	79
3.4.3.4 Michael Scriven’s Consumer Oriented Approach	80
3.4.4 Fourth Evaluation Wave	82
3.4.4.1 Robert Stake and Responsive Evaluation	83
3.4.4.2 Elliot Eisner’s Educational Connoisseurship Evaluation.....	84
3.4.4.3 House & Howe and the Deliberative Democratic Evaluation	85
3.4.4.4 Guba & Lincoln and Fourth Generation Evaluation (also known as Constructivist Evaluation Approach)	87
3.4.4.5 Patton’s Developmental Evaluation.....	99
3.4.4.6 Participatory, Empowerment and Collaborative Evaluation Approaches.....	100
3.5 Final considerations	103
Chapter 4 – Systems Thinking in Evaluation	104
4.1 Evaluation, systems thinking and complexity science. How are they related?.....	104
4.2 First wave of Systems Thinking and its application on evaluation.....	112
4.2.1 Systems Dynamics and Evaluation	112
4.2.2 The Viable System Model and Evaluation.....	116
4.2.3 Complexity Science and Evaluation	120
4.3 Second Systems Thinking Wave approaches in evaluation	123
4.4 Third Systems Thinking Wave approaches in evaluation.....	126
4.4.1 Critical Systems Heuristics and Evaluation	126
4.4.2 Systemic evaluation	128

4.4.3 Frameworks for evaluation practice.....	134
4.5 Final Remarks	137
Chapter 5 – Contribution	139
5.1 Reframing the Fourth Generation Methodology.....	142
5.1.1 The systemic character of Fourth Generation evaluation	142
5.1.2 Boundary judgments, boundary critique and the reframed form of Fourth Generation Evaluation	148
5.2 Consensus in Fourth Generation evaluation	149
5.3 Relevance.....	152
Chapter 6 – Contribution Methodological Design	153
6.1 Establishing a relationship with the client	156
6.2 Understanding the context in which the program and the evaluation take place	157
6.3 Identifying adequate stakeholders (the “who” component)	158
6.4 Reflecting upon the stakeholders’ selection.....	162
6.5 Developing, testing and enlarging constructions	163
6.6 Identifying and prioritizing relevant issues, claims, and concerns to be addressed.....	166
6.6.1 How to decide what to include and what to marginalize?.....	169
6.7 Reflecting upon the boundaries about “what” should be addressed by the evaluation.....	171
6.8 Identification and design of adequate tools for data collection.....	172
6.9 Data collection	173
6.10 Analysis of results.....	175
6.11 Discussing findings and generating recommendations.....	176
6.11.1 Discussion and sharing of results to generate recommendation with the stakeholders	176
6.11.2 Generating recommendations	177
6.11.3 Discussing findings and generating recommendations in practice	177
6.12 Producing reports	178
6.13 Recycling	178
6.14 Final considerations	178
Chapter 7 – The Hermes Program	181
7.1 Bogota Chamber of Commerce.....	182
7.1.1 Center for Arbitration and Conciliation	182
7.2 Alternative Dispute Resolution.....	183
7.3 The Hermes Program	184

7.3.1 Background	184
7.3.1.1 Legal Background	185
7.3.1.2 Historical Background	185
7.3.2 Objective	186
7.3.3 Methodology	187
7.3.4 Graduate Network	188
7.3.5 Operation of the Hermes Program	189
Chapter 8 – Hermes Program Evaluation	190
8.1 Establishing a relationship with the client	191
8.2 Organizing to conduct the evaluation	193
8.2.1 General context	193
8.2.2 Specific context.....	196
8.3 Identifying the stakeholders	199
8.4 Reflecting upon the stakeholders’ boundary	201
8.5 Developing, testing and enlarging constructions	204
8.5.1 Questionnaire based on Critical Systems Heuristics.....	207
8.5.2 Crazy trains	207
8.5.3 Drawing the community	208
8.5.4 Stories of Change.....	209
8.5.5 Six Thinking Hats	210
8.5.6 Final Debate and Sharing.....	211
8.6 Identifying and prioritizing relevant issues, claims, and concerns to be addressed.....	212
8.7 Reflecting upon the boundaries about “what” should be addressed by the evaluation.....	214
8.8 Identification and design of adequate tools for data collection.....	214
8.8.1 Students’ survey.....	216
8.8.2 Consultants’ survey.....	216
8.9 Data collection	217
8.10 Analysis of results.....	218
8.11 Sharing results and generating recommendations.....	218
8.12 Producing reports	218
8.13 Recycling	219
Chapter 9 – Hermes Program Evaluation Results.....	219

9.1 Workshop results	220
9.1.1 Activity 1: CSH based questions	220
9.1.1.1 Motivation.....	221
9.1.1.2 Commitment	223
9.1.1.3 Scope.....	227
9.1.1.4 PESTEL Variables	229
9.1.1.5 Impact of the program.....	231
9.1.1.6 Success variables	233
9.1.1.7 Other actors' perception about impact and improvement	235
9.1.1.8 Self-Reflection	237
9.1.1.9 Beneficiaries	239
9.1.1.10 Training.....	241
9.1.1.11 Additional Comments	242
9.1.2 Confidence Activity	244
9.1.3 Drawing the community	245
9.1.4 Small group activity	252
9.1.5 Six Thinking Hats	252
9.1.6 Overview of the results of the workshop	256
9.2 Student Surveys Results.....	257
9.2.1 Grade.....	258
9.2.2 Age.....	259
9.2.3 Gender.....	260
9.2.4 Distribution of students through the program phases	260
9.2.5 Household composition	261
9.2.6 Quality of the family relationships.....	262
9.2.7 How do you think is the perception of the Hermes Program in your school?	264
9.2.8 ¿How do you think people you live with perceive the program?.....	267
9.2.9 To which degree does the student agree with each of the statements presented in this question?	270
9.2.10 Frequency with which students are victims and promoters of different types of conflict	273
9.2.11 Student's behavior when witnessing conflicting situations between other individuals	278
9.2.12 Aspects the students like the most and the least about the Hermes Program.....	282
9.2.13 How do you think has been the implementation of the program in your school?.....	283

9.2.14 Correlation	288
9.3 Consultants survey results.....	289
9.3.1 Gender.....	290
9.3.2 Age.....	290
9.3.3 Level of Education: Which is your maximum completed level of education?	291
9.3.4 Field of studies: What is your career field?	292
9.3.5 Time as consultants of the Hermes Program: How long have you been working as a consultant of the Hermes Program?	293
9.3.6 Number of schools the consultant is working with.....	294
9.3.7 What do you think is done in the schools to improve the communication skills among the students that are participating in the Hermes Program?.....	294
9.3.8 What do you think is the degree of commitment of the different members of the community with the program?	295
9.3.9 What factor do you consider restrict the commitment of the different members of the school's community? What do you think can be done to increase their commitment?	297
9.3.10 How much do you think the relationships among the following actors have improved?	300
9.3.11 In your opinion which are the most important variables for a successful execution of the Hermes Program	301
9.3.12 Do you consider that the administrative requirements of the Chamber of Commerce (goals, indicators, paperwork, etc.) restrict the quality and the impact of the program?.....	303
9.3.13 Please justify your answer to the previous question	304
9.3.14 In your opinion, what is the importance of the Chamber of Commerce staff commitment in the success of the Hermes Program?	305
9.3.15 Question 15	306
9.3.16 Which activities do you think would motivate parents to get more actively involved in the program?	310
9.3.17 Are you satisfied with the current structure of the program?.....	311
9.3.18 Question 18	312
9.3.19 How do you perceive the general quality of the Hermes Program when comparing it with other programs focused on conflict resolution techniques?	313
9.3.20 Additional comments	314
9.3.21 Consultants Survey Correlation analysis	318
9.4 Findings.....	319
9.5 Conclusions about the Hermes Program Results	321
Chapter 10 – Power issues in the reframed form of Fourth Generation evaluation	325
10.1 Foucault on the exercise of power	327

10.2 Power and Knowledge	331
10.3 Power issues in the reframed form of Fourth Generation evaluation	333
10.4 Power-Knowledge and the reframed form of Fourth Generation evaluation.....	336
10.5 Final Reflections over power issues in the reframed form of Fourth Generation evaluation	341
Chapter 11 – Biases in the practice and development of the reframed form of Fourth Generation evaluation	345
11.1 Study of Biases	345
11.2 Sources of Bias	346
11.3 Forms of bias in Fourth Generation Evaluation Methodology	346
11.3.1 Common forms of bias.....	347
11.3.2 Other forms of bias	353
11.4 Biases and the reframed methodology of Fourth Generation Evaluation in Practice	354
11.5 Reflections over Fourth Generation evaluation and biases.....	357
Chapter 12 –Considerations about the Hermes Program Evaluation	359
12.1 Discussing Results and Generating Recommendations about the Hermes Program	359
12.2 Changes over the Hermes Program Evaluation	362
12.3 Power and Biases in the Hermes Program evaluation	366
Chapter 13 – Conclusions.....	368
References.....	379
Appendices.....	403

LIST OF FIGURES

Figure 1. The interdependence of boundary judgments, observations, and evaluations.	41
Figure 2. Critical Heuristic Categories.	42
Figure 3. Marginalization.	45
Figure 4. Margins, ethics, sacredness, profanity and ritual.	47
Figure 5. Model of overlapping stakeholder concerns, with attributions of profanity.	48
Figure 6. Three aspects of a methodology for systemic intervention. Taken from Systemic Intervention: philosophy, methodology and practice.	50
Figure 7. Evaluation Theory Tree (1st Edition)	55
Figure 8. Evaluation Theory Tree (2nd Edition).	56
Figure 9. Key components of the CIPP evaluation model and associated relationships with programs.	78
Figure 10. Relationship between the three approaches to evaluation.	132
Figure 11. A systems-based framework for rigor in evaluation.	137
Figure 12. Boundary critique in stakeholders' selection.	161
Figure 13. Boundaries of the issues, claims, and concerns to be tackled by the evaluation.	167
Figure 14. How to choose the “what” component of the evaluative process.	169
Figure 15. Prioritization of issues, claims, and concerns relevant for the stakeholders.	170
Figure 16. Steps for sharing findings and producing recommendations.	177
Figure 17. Methodological design of Fourth Generation evaluation by considering boundary critique and systems concepts.	179
Figure 18. The methodological structure of the reframed Fourth Generation evaluation approach.	180
Figure 19. The Hermes Program in Bogota.	189
Figure 20. Stage 1 Boundary.	202
Figure 21. Formed boundaries in stages 1 and 2.	202
Figure 22. Group 1 Community Drawing.	246
Figure 23. Group 2 Community Drawing.	247
Figure 24. Group 3 Community Drawing.	249
Figure 25. Group 4 Community Drawing.	250
Figure 26. Group 5 Community Drawing.	251
Figure 27. Grade Distribution.	258
Figure 28. Age Distribution.	259

Figure 29. Gender Distribution.	260
Figure 30. Distribution by program phase.....	260
Figure 31. Household composition.	262
Figure 32. Perceived Quality in the students' family relationships.....	262
Figure 33. Perceived quality of the students' family relationships according to their distribution throughout the program phases.....	263
Figure 34. Program Perception in the school.	264
Figure 35. Program perception in the school differentiated by the phase to which the surveyed students belong to.	265
Figure 36. Program perception at home.	267
Figure 37. Program perception at home considering the phases in which the students are.....	268
Figure 38. Mean of the students' perception regarding different issues related to the operation of the Hermes Program differentiating by the phase of the program in which the students are.	271
Figure 39. Mean about the perceived commitment of other stakeholders with the Hermes Program taking into account the program phase in which the students are located.	272
Figure 40. Frequency with which students see themselves as being victims of certain behaviors.	274
Figure 41. Frequency with which students have certain behaviors towards other students.	274
Figure 42. Mean with which students recognized themselves as being victims of certain behaviors taking into account gender.	276
Figure 43. Mean with which students recognized victimizing others differentiated by gender.....	277
Figure 44. Students' response when witnessing inappropriate behaviors among other students.	279
Figure 45. Percentage of students who used dialogue as a response to inappropriate behaviors differentiated by gender.	280
Figure 46. Frequency with which students chose to use dialogue in each one of the considered behaviors and differentiated by the phase in which the student could be in the program.	281
Figure 47. Students' preferences over issues related to the program.	282
Figure 48. Student's general perception about the Hermes Program.	284
Figure 49. Students' general perception of the Hermes Program differentiated by the program phases in which they could be.	285
Figure 50. Consultant's gender distribution.....	290
Figure 51. Consultant's age distribution.	291
Figure 52. Consultant's level of education.	291
Figure 53. Distribution of the consultant's fields of studies.....	293
Figure 54. Time as consultants.....	294

Figure 55. Consultant's perception about stakeholders' commitment with the Hermes Program.	296
Figure 56. Consultant's average perceived commitment of the stakeholders of the Hermes Program.	297
Figure 57. Impact of the Hermes Program over the relationships between stakeholders.	300
Figure 58. Frequency with which consultants consider certain variables to be determinant for the success of the Hermes Program.....	302
Figure 59. Impact of administrative requirements over the quality of the Hermes Program.	303
Figure 60. Importance of the commitment of the Chamber of Commerce over the success of the Hermes Program.	306
Figure 61. Likelihood of the consultants of working with the Hermes Program in 2017.	307
Figure 62. Likelihood of the consultant for working with the Hermes Program in the long run. ...	307
Figure 63. Perception over the impact of the consultant's rotation over the performance of the Hermes Program.	308
Figure 64. Consultant's perception over the possibility of working more time with a narrower number of schools.....	309
Figure 65. Frequency with which possible ideas for fostering parental motivation are chosen. ...	310
Figure 66. Consultants' satisfaction with the current structure of the Hermes Program.	312
Figure 67. Consultants' general perception about the program.	314

LIST OF TABLES

Table 1. F-test for testing differences in the perception of the program at home given the students' phases.....	264
Table 2.F-test for testing mean differences between the students' perceptions about the perception of the program in the schools.....	266
Table 3. Tukey test for mean differences with regards to the program perception in the schools.	266
Table 4. F-test for the differences in the mean of the perception of the Hermes Program at the students' homes.	269
Table 5. Tukey test for the perception of the Hermes Program at home for each one of the phases in which students could be located.	269
Table 6. T-test for differences in the mean of the administrative staff and the teachers' commitment.....	273
Table 7. T-test for the mean with which males and females mock at others.	278
Table 8. F-test for the use of dialogue among the students in the different program phases.....	282
Table 9. F-test for the program perception for the students in each one of the program phases.	286
Table 10. Tukey test for the program's perception among the students in its different phases.	287
Table 11. Correlation between the students' survey variables.	288
Table 12. Number of schools per consultant.....	294
Table 13. Hermes Program quality perception.	314
Table 14. Consultant's survey correlation analysis.	318

LIST OF ABBREVIATIONS

ADR	Alternative Dispute Resolution
BCC	Bogota Chamber of Commerce
CAC	Center for Arbitration and Conciliation
CAS	Complex Adaptive Systems
CATWOE	Customers, Actors, Transformation process, Weltanschauung, Owners, Environmental constraints
CIPP	Context, Input, Process, and Product
CLD	Causal Loop Diagram
CSH	Critical Systems Heuristics
CST	Critical Systems Thinking
DSRP	Distinctions, System, Relationships, Perspectives
ECLAC	Economic Commission for Latin America and the Caribbean
GST	General Systems Theory
HSD	Human Systems Dynamics
LSI	Local Systems Intervention
MCE	Model for Collaborative Evaluation
NGO	Non-Governmental Organization
PAR	Participatory Action Research
PESTEL	Political, Economic, Social, Technology, Environmental, and Legal variables

SAST	Strategic Assumption Surfacing and Test
SD	Systems Dynamics
SSM	Soft Systems Methodology
SoSM	System of Systems Methodology
STCS	Systems Thinking and Complexity Science
UTOS	Units, Treatments, Observations and Settings
TSM	Total Systems Intervention
VSM	Viable System Model

Chapter 1 – Introduction

1.1 General Aim of the Thesis

The general aim of this thesis is to explore how systems thinking concepts could be helpful for conducting social program evaluations. In order to do so, a reframing of a program evaluation approach or model called Fourth Generation evaluation was conducted by introducing systems thinking concepts (particularly those related to boundary critique and boundary judgments), power issues, and the research on biases associated with human judgment in it. Further, the proposed methodology is used in evaluating the Hermes Program, a program created by the Bogota Chamber of Commerce with the aim of teaching children how to solve conflicts by means of alternative conflict resolution techniques.

1.2 Specific Aims of the Thesis

The specific aims of the research are presented in what follows. They are divided in larger aims supported in sub-aims by means of which they are achieved.

1. To design, develop and put into action a program evaluation methodology based on the inclusion of systems thinking concepts. Such evaluation approach could be adapted to be used in evaluating social programs involving the participation of multiple stakeholders by taking into account the contextual characteristics surrounding the evaluand, the evaluation and the particular needs that must be fulfilled. This aim involved other sub-aims as follows:

1.1 Reviewing and choosing the most appropriate systems thinking concepts to be included in a program evaluation approach.

1.2 Exploring the work that has been done so far in introducing systems thinking concepts in program evaluation at both, theory and practice.

1.3 Reviewing and selecting a program evaluation approach (or model) over which systems thinking concepts will be cast upon, an approach which may also be adequate for conducting the practical case.

1.4 Reframing a program evaluation model by taking into account the chosen systems thinking concepts.

2. To understand how power issues may emerge and affect the theory and practice of the proposed methodological design. This involves:

2.1 Exploring the conceptions of power over which program evaluation is traditionally

grounded and particularly the foundations over which the chosen evaluation approach is set.

2.2 Framing power issues in the chosen evaluation approach in terms of Foucault's conception of power.

2.3 Understand how different power conceptions affect the methodological evaluation development and how this reflects in practice.

3. To explore how cognitive biases may be present in the proposed methodological development and to propose mechanisms by which they could be addressed.

3.1 Review what has been done so far in terms of biases within the program evaluation field.

3.2 Recognize which are the most relevant cognitive biases for evaluation at both; theory and practice.

3.3 Describe how biases may be present at a theoretical level in the proposed evaluation methodology and how such biases may be handled by the methodological steps and the practice of the evaluation.

4. To develop a new narrative over how the evaluation field has developed so far. This requires:

4.1 Reviewing different mapping methods developed throughout the evaluation field.

4.2 Recognizing a differentiating factor between my mapping design and those developed by other authors.

4.3 Recognizing and choosing common traits and patterns over the mapping methods, as well as over the program evaluation approaches.

4.4 Identifying the issues and critiques that led to further developments in evaluation.

5. To conduct the evaluation of a program developed by the Bogota Chamber of Commerce by using the proposed methodology development. This involves:

5.1 Learning what the Hermes Program does, how it operates, and the contextual factors surrounding it.

5.2 Adapting the proposed methodological design to the context and needs in which it

will be applied.

5.3 Choosing and designing data collection tools as needed.

5.4 Analyzing results and producing recommendations for the operation of the Hermes Program.

1.3 Relevance

This research has both, a theoretical and a practical relevance that will be described in what follows.

1.3.1 Theoretical Relevance

In the American Evaluation Association Conference held in 2017, Michael Quinn Patton along with other panelists, addressed the need for leaders and evaluators of developing a critical evaluation thinking in order to achieve the United Nations Sustainable Development Goals (SDG). By these means, they recognized the necessity of thinking outside the box, the need of joining leadership and evaluation, and the demand for capacity building in trading for evaluation building across the globe by posing the following question: How do you use evaluation to improve the life of people? In responding such question, they realized the importance of interconnectedness of the SDG's. In this discussion, Patton identified 5 evaluative leadership qualities that should be reinforced: think beyond projects, think beyond nation-state borders, think beyond SDGs siloes, to move beyond top-down vs bottom-up to dynamic local-global interconnections, and to move from the independence to interdependences. This represents the need of a transformational change in leadership and evaluation in which systems thinking and particularly critical systems thinking can play a fundamental role to develop the current evaluation agenda.

However, when exploring what has actually been done in introducing systems thinking concepts in evaluation, and particularly in program evaluation, one can see that work still needs to be done. Although there has been interest within the evaluation field in exploring the use of systems thinking concepts at a theoretical level (Cabrera, Colossi & Lobdell, 2008), most of the work that has been conducted in introducing such concepts has been directed towards using systems thinking developments as methods for supporting practical cases of evaluation (e.g Burke, 2006; Dick, 2000; Sharma & Gutierrez, 2004; Williams & Hummelbrunner, 2011) and not to alter the methodological foundations of doing so. Furthermore, although several evaluators claim to be using or to have used systems thinking

and complexity (independently) ideas and methods in evaluation, the extent and ways in which evaluators are drawing on these ideas is not well understood or operationalized (Gates, 2016; Reynolds et al, 2012; Walton, 2016). To date, there has been no systematic examination of how these ideas and methods contribute to the [program] evaluation field or a framework for understanding when and why to use them (Gates, 2016). Moreover, the way in which evaluation theory and practice operate is a concern as well. Recent research on evaluation illustrates that, despite many noteworthy efforts to integrate evaluation's theory into its practice and its practice into its theory, more often than not, the gap between theory and practice remains a chasm (Christie, 2003; Christie, Quinones, & Fierro, 2014; Coryn, Noakes, Westine, & Schroëter, 2011; Miller & Campbell, 2006).

This suggests the presence of a gap regarding the use of systems thinking in the evaluation field, not because attempts of producing such inclusion have not been done but because they are perceived as incomplete or not completely clear. There is a need then of not only incorporating systems thinking at a theoretical level in evaluation but also on detailing how such amalgamation works, why is it useful and how they will both work in practice. In doing so, one can also identify a second gap related particularly to the connection between the theory and practice of the evaluation field. It seems like although there have been substantial advancements in evaluation theory, such advancements do not always translate into practical advancements, as they are not regularly used. The research described in this document does not only represents an attempt of including systems thinking concepts in evaluation but also a way in which theory can be put into practice by means of using the proposed methodology in conducting a real life evaluation.

As I have already stated, incorporating systems thinking in program evaluation is an uncompleted task. However, the work of Cabrera, Colossi & Lobdell (2008) and Imam, LaGoy & Williams (2006) represent the most important work in introducing systems concepts in evaluation that has been conducted so far. Both groups of authors recognize interrelationships, perspectives, and boundaries as essential concepts for conducting a systemic evaluation, although Cabrera, Colossi & Lobdell (2008) also recognize the system itself as a fourth relevant concept. Despite of representing an advancement in introducing systems concepts in evaluation, both works limit to proposing the identified concepts without exploring how they may be useful or relevant at a theoretical or practical level. Moreover, Patton's (2011) work on Developmental Evaluation constitutes the only attempt in the evaluation field of introducing complexity concepts (not systems thinking) in a particular

program evaluation approach.

As a result of these considerations, one can find a theoretical gap not only related to how and which systems thinking concepts are introduced in the evaluation field as a whole, but also on how such concepts could be included in particular evaluation models and approaches. This research intends to fill this gap by focusing on reframing and introducing systems thinking concepts in the methodology of a program evaluation model.

The following statements show some of the benefits of including systems thinking concepts in program evaluation:

1. Evaluation has explored merit and worth, processes and outcomes, formative and summative evaluations. The great unexplored frontier is evaluation under conditions of complexity (Patton, 2011).
2. Just as interventions are faced with increasing complexity, so are the evaluators of these interventions. They are often confronted with a diverse or even conflicting picture of reality; for example, when seen through the eyes of various stakeholders or compared to some pre-defined (mostly linear) logic. And attempts at reducing this in an inappropriate manner can lead to questionable or even misleading findings, harm the relevance of the evaluation and bring forth resistance from those which feel not properly represented (Hummelbrunner, 2011).
3. Systems thinking can help contribute towards a shift from 'attribution' (purposive systems) to 'contribution' (purposeful systems design) and 'beyond' (critical thinking on the ethics and politics of evaluation) (Reynolds, Forss, Hummelbrunner, Marra , & Perrin , 2012).
4. Systems concepts will be particularly useful to evaluators in situations where rigorous rethinking, reframing, and unpacking complex realities and assumptions are required (Imam, LaGoy, & Williams, 2006).
5. Numerous developments in evaluation theory and methods, including Theories of Change approaches, Realist Evaluation and Fourth Generation Evaluation, have created a climate that is permissive to use complexity thinking within evaluation (Walton , 2016).

Another important issue regarding evaluation is power. This is a relevant issue given that evaluation is in itself a political act. Evaluation involves politics at both the micro and institutional levels, and in multiple ways, since it engages complex social relations and

decisions about rules and resources between stakeholders, or different interest groups, that have vested interests in the outcomes of the evaluation (Legge, 1984; Patton, 1981). Even before an evaluation begins, the act of deciding whether something should be evaluated is a political act (Taylor & Balloch, 2005). Morris (2011), states that the American Journal of Evaluation (AJE) authors have approached the notion of politics as the interactions of various stakeholders in evaluation as they articulate their interests from different positions of power, influence, and authority. During the evaluation itself, the consequences of the evaluator's actions can affect downstream stakeholders not immediately present. Thus, more so than other types of research, the environment of evaluations can lead to political situations (O'Brien, Payne, Nolan, & Ingleton, 2010). Although the political connotation of evaluation has been comprehended by exploring what forms of knowledge are perceived as more valid and which structures of power determine them, power is generally understood as something infringed over others, which can be possessed in different extend by the stakeholders (Nolan *et al.*, 2007; O'Brien *et al.*, 2010; van Draanen, 2016).

There are multiple standing points from which power could be understood, for this reason an identified gap would be associated to the possibility of exploring power from a standing point from which it is not attributed a negative or a positive character but a productive one. Such conception would be different from the traditional understanding of power in the program evaluation field. Reframing the traditional understanding of power, implies changes in its possession character. In order to do so, I seek to frame my own systemic methodological evaluation design in terms of understanding power under Michel Foucault's ideas (1979, 1990) and studying which are the implications of doing so at a theoretical and a practical level.

At the same time, I also seek to explore how biases are present in the theory and practice of evaluation. Up to what I know, biases in evaluation have only been explored by Scriven (1975). However, Scriven addresses biases as prejudices that undermine the objectivity with which an evaluation is conducted, leading to invalidity, lack of reliability, reduced credibility, and spurious outcomes. Under this view, it becomes fundamental to remove at any cost biases from evaluation, otherwise the evaluative process and its results may be considered as unreliable. This turns out to be important given that objectivity is understood on the basis of universality. Nonetheless, I would argue that although one can produce measures to handle biases, they cannot be completely eliminated from situations implying decision making processes. Up to what I know, aside of Scriven's exploration of biases in evaluation, further

work has not been conducted on this issue in the evaluation field, even less with regards to Fourth Generation evaluation, which is the specific evaluation methodology addressed in the current research. For this reason, this analysis is also a way of filling a gap which has not so far been recognized in the field. This gap is driven by the following questions: How do biases influence evaluation at a theoretical and practical levels? How can they be tackled? In addressing such questions one can explore which may be relevant biases for evaluation, not only in formulating a methodological design but also in putting such methodology into practice. The analysis of biases in the current research involves three components; understanding how biases are formed, comprehending how biases may emerge at a methodological level, and describing how they were tackled in practice.

1.3.2 Practical Relevance

Aside of having a theoretical relevance, this research has a practical relevance as well.

First, I have stated that there exists a gap between the theoretical developments and the practical cases of program evaluation. For this reason, this research constitutes an opportunity for showing how particular evaluation methodologies are put into practice, and how such evaluations are adapted to the contextual characteristics surrounding not only the evaluation but the evaluand as well.

Second, given that the practical case of the evaluation is directed towards evaluating a program developed by the Bogota Chamber of Commerce, the results of the evaluation are expected to be useful for them in addressing the performance of the program and how improvement measurements could be proposed. This in extend could contribute to the operation of the program and how stakeholders participate in it and its evaluation the future.

Third, the methodology followed to evaluate the program could be useful for evaluating other social programs involving the participation of different stakeholders, as well as for making future evaluations of the Bogota Chamber of Commerce social programs.

Fourth, this research could be useful not only for addressing how evaluation theoretical methodologies are put into practice but also on how systems concepts are included in evaluation not only in theory but how such inclusion takes place in practice. This is important, given that one of the difficulties attributed to the attempts that have been made so far in including systems thinking concepts in evaluation is the fact that there does not exist any practical case that shows how this has been done.

1.4 Structure of the Thesis

Chapter 1 of this document, which you are reading right now, corresponds to the introduction of the research. It is directed towards presenting the aims with which the research was conducted, as well as its relevance at a theoretical and a practical level. Following this chapter, comes *Section 1*. This section comprises chapters 2 to 4 which are directed towards presenting the theoretical review of the research.

Chapter 2 describes the historical evolution of the systems thinking field. In order to do so, I first describe the emergence of the system concept and then I move on to describing the progress of the field by using waves. Each one of the waves is presented by outlining their core ideas, those which in my opinion constitute their most significant developments, and the criticism produced with regards to each one of them, which ultimately lead to the emergence of new waves. Special attention was placed in describing the development of boundary related concepts such as boundary critique and boundary judgments as they are a central element in the development of this research.

Chapter 3 is oriented towards presenting the development of the evaluation field. In it, I first focus on identifying different conceptions of the evaluation term and selecting the one I consider as more appropriate for grounding my research. In giving an account of the evolution of the field, I start by presenting two ways in which other authors developed their own evaluation theory upon which my own development is grounded. Afterwards, I proceed to present my own development for describing the progress of the evaluation field. Once again I decided to use a waved approach by which I present the most important evaluation models that belong to each one of them and their critique. Given that the research is focused in working with Fourth Generation evaluation, particular attention and detail is given to this model.

Chapter 4 is the chapter in which the systems thinking and the evaluation fields are intertwined. In this chapter I present how systems thinking has been brought to the evaluation field in both theory and practice. Following the same structure of the two previous chapters, this chapter is divided according to how developments within the three systems thinking waves have been used in evaluation. As some of the practical uses of systems thinking in evaluation do not fall in a particular wave, there is a particular section dedicated to addressing their application in a separate way.

Section 2 comprises chapters 5 and 6 both of which are directed towards introducing my own

methodological design. Chapter 5 outlines how systems thinking concepts are introduced in the methodology of Fourth Generation evaluation. I start by addressing the reasons for which the evaluation model or approach as well as the systems thinking concepts were chosen. Once this has been done, I address how this form of evaluation acquired a systemic character by the introduction of these concepts as well as addressing how each one of them is introduced in the evaluation methodology. The chapter ends by considering the relevance of introducing these concepts in the practice of evaluation.

Chapter 6 constitutes a detailed description of what was described in Chapter 5. This chapter is focused on presenting the original methodological steps of Fourth Generation evaluation, and on addressing the reframed methodology one step at a time by describing how new steps were introduced, other steps were altered, and how specific steps of the methodology are affected by following a rationality grounded in systems thinking.

Once my methodology has been introduced it is time to present how it was put into practice. For this reason, *Section 3*, comprising chapters 7 to 9, refers to the Hermes Program evaluation. As the reader is not familiarized with the Hermes Program, Chapter 7 introduces it by means of an outline including its historical background, objectives, areas of operation, and methodology.

Following this introduction, Chapter 8 focuses on describing in detail how the Hermes Program evaluation took place. Such recount, is presented following the stages of the reframed Fourth Generation methodology by describing the activities conducted in each one of the stages. Some of the stages are only outlines as they refer to the production of some results that will be addressed more extensively in the next chapter.

Chapter 9, is focused on presenting the results of the evaluation of the Hermes Program. The reasons for addressing results independently from the other stages of the evaluation is that first, more than one stage of the evaluation produced results for which it was easier to address them jointly. Second, results were obtained by means of a wide variety of means described in the previous chapter, for which they were analyzed using different means and tools. The chapter starts by describing the results of a workshop conducted with the stakeholders of the program, for which all the results were qualitative. The second part of the chapter, presents data analysis from two groups of surveys using STATA as the statistical tool by which data was analyzed. The final part of the chapter presents a general conclusion of the findings of the evaluation results.

Section 4 comprises chapters 10 to 13. This section explores the reframed form of Fourth Generation evaluation from a different point of view involving issues related to biases and power. Such exploration, took place after the methodology was reframed and put into practice. This section also present the considerations of the methodology in practice and the final conclusions of the research. Chapter 10 explores the conception of power under the proposed view of Michel Foucault. I start this chapter by presenting Foucault's ideas on power and power-knowledge, ideas which I consider fundamental for the theory and practice of the reframed form of Fourth Generation evaluation. Once such conceptions have been outlined, I illustrate how these concepts affect the methodological conception of the reframed form of Fourth Generation evaluation as well as how they oppose to the initial conception upon which the methodology was grounded.

Chapter 11 addresses issues on biases. In order to tackle this, I start by introducing how biases emerge at a cognitive level. As approximately 100 types of cognitive biases have been recognized by researchers, in this chapter I only address the ones I consider as having the possibility of emerging in the reframed form of Fourth Generation evaluation. In doing so, I briefly define each bias and how it works, as well as the way it may appear at a theoretical level in the methodology. Afterwards, I move forward to explore how in the Hermes Program evaluation some of these biases were tackled as well as proposing ways in which they could be handled in practice. The chapter ends with general reflections over the reframed form of Fourth Generation evaluation and biases.

In Chapter 12, I present some considerations regarding the Hermes Program evaluation. Such considerations have to do with how activities that were not conducted could have been carried out, a reflection over how evaluation step that actually took place could be done differently, as well as a reflection over how biases and power issues may be considered in the evaluation of the Hermes Program.

Finally, Chapter 13 describes how the aims of the research outlined in the introduction were fulfilled.

Chapter 2 – Systems Thinking

Throughout the twentieth century, the development of systems thinking was characterized by shifts in paradigms expressed in different waves of research (Midgley, 2003). This chapter seeks to address how these shifts took place. I will start by exploring the emergence of the system concept and the General Systems Theory, a theory which set the ground for the development of the first wave of systems thinking as it incorporated key concepts that were not considered in the past, such as the open system. After describing the General Systems Theory I will move on to describing the development of each one of the systems thinking waves, focusing primarily in the developments that took place in the third wave, where my own research is located. In order to do this, I will not only describe the key concepts underlying each wave but I will also focus on the critique towards each one of them, which eventually led to reframing and expanding the work that had been previously done.

2. 1 Precedent to Systems Thinking

Historically there have been several developments that lead to the establishment of Systems Thinking, in the following section I will describe two of the key developments in the field of systems science in the twentieth century. Systems science is an integrative discipline which brings together ideas from a wide range of sources which share a common systems theme (SEBoK authors, 2017). I will start by describing the work of Angyal who was the first author to coin the term ‘system’ (at least in the way used by modern systems thinkers) and the work of Von Bertalanffy and Boulding which was consolidated through the development of a General Systems Theory. General Systems Theory, is seen as the ground for the development of systems thinking, mainly because it led to a change in the conception of organizations and management sciences.

2.1.1 The emergence of the system concept

The modern system concept was arguably first introduced by Angyal (1941) as a logical “genus”, this means a class of things that have common characteristics (Oxford Dictionary, 2017), for the treatment of wholes, under the ideal of developing a logic of system to a degree of precision that it might offer the basis for exact mathematical formulation of holistic connexions. Holism refers to the theory in which parts of a whole are in intimate interconnection, such that they cannot exist independently of the whole, or cannot be

understood without reference to the whole, which is thus regarded as greater than the sum of its parts (Oxford, 2016). This idea was developed as a counter position of the general belief that the structure of wholes could not be dealt with through a logical manipulation and as an answer to the problem of integrating part processes in the total organization of living systems.

The author clarifies that the way in which he uses the term “system” differs from the way it is usually understood. According to Angyal (1941), the traditional conception of system is used to label an aggregate of elements considered together with relationships holding among them. However, here the term system is used to denote a holistic system, this is the abstraction of its constituent elements and referring only to the organization of the whole. Thus, it is a holistic organization. From the consideration of this definition, the author proposed a set of characteristics that differentiate a simple relationship from a system.

First, whereas a relationship implies a relation between two sides, a system may involve an unspecified number of members. Second, a relation requires an aspect out of which the relationship is formed, being it an attribute or property of the object (colour, size, shape etc.). However, those elements that are part of a system, are not related through their immanent qualities but by their arrangement within the system. The third characteristic of the system is that the dimensional domain is more involved in the formation of systems than it is in the consideration of relationships between objects, as a system is a distribution of its members in a dimensional domain. Finally, Angyal (1941) describes the complexity of the connexion of the members of a system. In a system the members are, from a holistic viewpoint, not significantly connected with each other except with reference to the whole. The constituent parts of a system are not considered separately but with respect to a superordinate, more inclusive factor, the system in and by which they are connected (Angyal, 1941). This last characteristic is an essential part of the author’s conception of system.

We will now describe the development of the General Systems Theory, which took place few years after the proposal of the system concept given by Angyal. However, the two conceptions share several characteristics, being the main one the consideration of the whole over the consideration of the relationships of the parts.

2.1.2 General Systems Theory

During the first part of the twentieth century, the dominant world view was a mechanistic one, implying an analytical reductionist science. This type of science continually divided its subject matter into more and more specialist disciplines, focusing on a small number of

unilinear causal relationships between phenomena and looking for explanations of all phenomena in terms of their smallest identifiable parts (Midgley , 2003). However, around the 1950s several authors started to realise that this worldview was not adequate to handle the emergence of concepts, viewpoints and conceptions with a similar character that were taking place in diverse fields of science. This is how General Systems Theory (GST) emerged, as a discipline that recognized the necessity of studying things as a system. That is, recognizing the necessity of understanding and studying things considering the organizing relations that result from dynamic interactions and make the behaviour of parts different when studied in isolation or within the whole (Von Bertalanffy, 1956). This organizing relations are key elements of GST as they were to Angyal (1941), as they give rise to properties that cannot be attributed to things viewed in a separate way in which the whole is more than the sum of its parts.

GST assumes the existence of general common system properties and characteristics in different fields, this is coined as “isomorphies”. For this reason, its main goal is to formulate and derivate principles that are valid for systems in general (Von Bertalanffy, 1956). That is, the development of unified principles running “vertically” through the universes of the individual sciences, bringing us nearer to the goal of the unity of science (Von Bertalanffy, 1956). GST describes the common characteristics of systems by using a mathematical and an ordinary language approach, although they recognize that this is not the only way of doing so.

It is important to clarify that although the unity of science might be the ultimate goal of GST, it does not seek to establish a single, self-contained “general theory of practically everything” which will replace all the special theories of particular disciplines (Boulding, 1956). Instead, the unitary principle is that we find organization on all levels, and this conception being based on the isomorphy of laws in different fields. These two authors, represent the work of GST.

Although Von Bertalanffy’s (1956) and Boulding’s (1956) work on GST is directed towards the conception of unity of science, both authors developed GST in different ways. Von Bertalanffy’s work was directed more towards the development of theoretical models for achieving integration while being able to handle concepts such as wholeness, directiveness, teleology, control, self-regulation, and differentiation which depict the need for interaction. Whereas the work of Boulding was directed towards arranging the empirical field in a hierarchical complexity of organizations of their basic “individual” or unit of behaviour, and

to try to develop a level of abstraction appropriate to each (Boulding, 1956). This approach is an arrangement of theoretical systems and constructs in a hierarchy which not only sought to map theories but also to be able to show how to fill the gaps between them. In order to do so, the hierarchy includes eleven levels of theoretical discourse, starting with a static structure (the level of frameworks), up to reaching the level of transcendental systems. Although both authors developed GST in different ways, this does not mean that their developments have a competitive character, instead they have a complementary one (Boulding, 1956).

One of the key elements of the development of GST and its view of systems is the relationship between the system and its environment. The reductionist science, which had been dominant until the moment, dealt only with closed systems, these are systems that are isolated from their environment. However, with GST authors realized the need of a shift from considering closed systems to open systems. Open systems maintain themselves in a continuous inflow and outflow, building up and breaking down of components, never being, so long as it is alive, in a state of chemical and thermodynamic equilibrium but maintained in a so-called steady state which is distant from the latter (Von Bertalanffy, 1950, 1956). In simpler terms an open system is one that does have a continuous interaction with its environment, this means that it has inflows (from the environment) and outflows (to the environment), being in permanent change but being in a steady state as well, mainly due to its interchanges with the environment surrounding it. This represented a shift not only for opening systems but also because in the mechanistic view, interaction was viewed as being one way, but in this new conception system elements are seen as having a mutual interaction.

As can be seen, the development of GST brought to the fore a completely new conception of system and most importantly introduced the concept of organization in the definition of a system, as well as viewing a system as being an open one. From these developments, people started looking organizations as systems, leading to the development of the field of Management Systems and the different waves of Systems Thinking.

2.2 The First Wave of Systems Thinking

Before continuing with the description of the systems thinking waves, it is important to make clear that the work of Jackson (1991b) and Midgley (2000) has been considered as a guide for the development of the historical description of the three waves of systems thinking in this document. However, it should be considered that although the historical development here described is based on the ones presented by both authors, it does not exactly fit with the

historical developments described by them. Instead, it seeks to depict the most important developments in the view of the writer.

2.2.1 Development of the first wave

The first wave of systems thinking took place in the late 1950's up to the early 1970's, grounding its work in systems ideas that were being developed at the time, one of which was a key concept of GST, namely: the open systems theory of Von Bertalanffy (1950). The development of this concept had as a major consequence, a shift on the way organizations were seen, as the view moved towards viewing organizations as systems and in that matter, the development of management systems took place.

Several developments took place during the first wave of systems thinking, however in what follows I will be focusing on those I believe are the most relevant ones for the current research. These developments correspond to Socio-technical Systems Thinking and the Systemic Operational Research. The latter comprising Systems Analysis, Systems Engineering, Viable System Model (VSM) and Systems Dynamics. I will briefly describe each one of these approaches.

The first one of these developments was the Socio-Technical Systems Thinking, this approach was developed from the research work of the Tavistock Institute of Human Relations in the British Coal Mining Industry in the 1950's and the early 1960's (Trist & Bamforth; 1951, Emery, 1959; Trist., *et al*, 1963; Emery & Thorsrud, 1976). It emerged as an alternative to the technocratic bureaucracy in which self-regulating, multi-skilled work groups become building blocks for a more democratic and efficient organizational form. In doing so, it rejected the technological imperative and instead sought to find the best match between, or, in systems language, the joint optimization of, the social and technical system (Trist, Higgin , Murray , & Pollock, 1963). This approach had two key elements at the core of its development, first the importance of semi-autonomous working groups and second the recognition of social and technical elements of an organization (Midgley , 2000). Emery & Thorsrud (1976) recognized that individuals have limitations as building blocks in organizational design mainly due to technological reasons as the control of processes and information handling cannot be dealt with entirely on an individual basis. For this reason, semi-autonomous working groups involves the establishment of work groups with specific responsibilities and with conditional autonomy to undertake the task in a way to be decided within the group but constrained by the conditions within which it works (Susman, 1976).

The recognition of the technical and social elements of an organization are the other key aspect of socio-technical systems thinking, this approach not only recognizes these elements as being equally important but it also seeks to optimize them in a joint manner, as can be seen in the definition given above. As Jackson (1991b) states, the attainment of optimum conditions for any of the elements may not result in optimum performance for the system as a whole.

Moving on to the developments comprised by the Systemic Operational Research, we first find Systems Engineering. In order to describe what Systems Engineering is about, I first have to define the system concept under this view. A system is any group of human beings and machines with a definite objective (Jenkins & Youle, 1968). In terms of these definition Jenkins & Youle (1968), define systems engineering as the study of complex systems in their totality so that the component systems can be designed and fitted together so as to achieve the overall objective in the most efficient way. According to Jenkins (1972), the purpose of systems engineering is to ensure the optimal use of resources (human, material, technological, etc.). This can be achieved by a methodology incorporating four basic phases: systems analysis, systems design, implementation and operation.

On the other hand, Systems Analysis can be understood as a systemic approach for helping a decision-maker choose a course of action by investigating his full problem, searching out objectives and alternatives, and comparing them in the light of their consequences, using an appropriate framework – in so far as possible analytic – to bring expert judgment and intuition to bear on the problem (Quade & Boucher , 1968). It has both, a quantitative and a qualitative side as it not only encompasses a problem as such but also all the factors and concepts surrounding it, in terms of Miser & Quade (1985:2), the knowledge and methods of modern science and technology, in combination with concepts of social goals and equities, elements of judgment and taste, and appropriate consideration to the larger context and uncertainties that inevitably attend such system. In a practical level, Systems Analysis has been widely used in addressing issues related to science and engineering for industrial or military applications and with budgetary decisions where it points towards the most cost-effective courses of action (Hordijk, 2007).

Another development of Systemic Operational Research was System Dynamics, which constitutes an attempt to provide a common framework for the separate functional areas of management activity. In order to do so, Forrester (1962) developed a methodology for

building mathematical models of industries and a computer program to compute the complex interaction of the models. By varying inputs on the computer runs, the technique is supposed to reveal critical system variables that can then be manipulated to achieve better results in the management of an enterprise. Sterman (2001), describes it as a perspective that set of conceptual tools that enable us to understand the structure and dynamics of complex systems, to build formal computer simulations of these complex systems, and use them to design more effective policies and organizations.

The last of the developments of the Systemic Operational Research is the Viable System Model (VSM) which was developed by Stafford Beer in 1972 by coupling his understanding of open systems represented by the work of Von Bertalanffy (previously described) with the cybernetic system concepts of information and regulation through feedback. VSM was developed to demonstrate that certain common features are necessary to any system's viability or survival (Beer, 1985). Viability is a system's ability to maintain a separate existence. To accomplish this, a system must have the capacity to respond to the complexity of its environment and environmental change, even if the change was not foreseen at the time the system was "designed" (Brockelsby , Cummings , & Davies , 1995). Beer (1972, 1985) conceptualizes all viable systems as sharing five complementary subsystems defined as operational elements that exist to undertake the systems basic activities and they comprise: coordination, control/audit, intelligence, and policy functions.

According to Checkland (1985), the nature of this tradition can be summed up as follows: "it seeks to make possible the efficient achievement of goals and objectives, taking goal seeking to be an adequate model of human behaviour; it assumes that the world contains systems which can be "engineered", hence the models of those systems can be made; it talks the language of "problems" and "solutions" which eliminate problems". We will now describe the critiques directed towards the developments in the first wave of systems thinking.

2.2.2 Criticism of the First Wave of Systems Thinking

Criticism regarding the first wave of systems thinking was mainly developed by authors such as Churchman (1970), Ackoff (1979) and others who were once founding figures of the goal-seeking conception of purposeful systems. This criticism was mainly directed towards the developments of Systemic Operational Research.

Churchman (1970) recognizes several problems and gaps within the first wave of systems methodologies practices. He argues that practitioners fail to recognize the systemic

assumptions underlying data collection, as they think about this process as being the result of observation alone. This systemic judgments implicit in data includes judgments about the nature of the decision maker as well as the boundaries of the system. He argues, that data collection involves strategic decisions, i.e. alternative choices which have to be evaluated in terms of desired outcomes.

In the first wave of systems thinking, decision makers tend to understand the systems they are controlling into its component parts and not as wholes, this is mainly a result of the use of models. Regarding this subject, de Geus (1992) argues that no model can ever be a precise representation of reality. In the living world, like in the material world of inanimate objects, the observation of reality is influenced by the position of the observer (de Geus, 1992). This is directly connected to the criticism of Churchman towards the underlying assumptions of data collection, which ultimately means that models and data do not have an independent character. Jackson (1991b) states that this is a clear example of systemic modernism as it rests upon a “predict and prepare” paradigm in which history is unilinear and the future is susceptible to forecast. This type of relations, single simple linear, are no longer useful in a world that is focused not only in processes but also in the human relations that take place in real life where artefacts that communicate, feel and think are present.

Churchman (1970) and Ackoff (1979) are also deeply concerned with the value given to means and ends in the first wave practices. Here, the value of a mean is taken to be instrumental, extrinsic as its value lies on its ability of achieving certain goals. The value of an end is taken to lie in the satisfaction its attainment brings, to be purely non-instrumental intrinsic. It does not acknowledge, let alone take into account, the intrinsic value of means and the extrinsic value of ends, this is, every mean has intrinsic value because it is also an end, and every end has extrinsic value because it is also a means (Ackoff, 1979; Ulrich 1983; Flood & Ulrich, 1990).

Criticism is also directed towards the objective character of the first wave practices. Practitioners, mainly in the Systemic Operational Research branch, are always seeking “objective” results as they seem to arise from clearly warranted observation and reason. The prevailing concept of objectivity is based on a distinction between an ethical-moral man, who is believed to be emotional, involved and biased, and a scientific man, who is believed to be unemotional, uninvolved and unbiased (Ackoff, 1979). This dichotomy is the main reason why it does not incorporate the arts and humanities, as the belief of doing so would reduce its

objectivity, making the results less reliable. What practitioners do not realize is that results also depend on strongly systemic judgments. There is no such thing as a completely objective result mainly because what may seem an improvement for someone will be seen as a setback by other individual or group. This inability to incorporate other fields of knowledge lead to the first wave developments, mainly the systemic operational research ones, to become unidisciplinary without being capable of handling real life messes, which are more than a deconstruction of simple parts. As Ackoff (1979) states, objectivity is not the absence of value judgments in purposeful behaviour. It is the social product of an open interaction of a wide variety of subjective value judgments. Objectivity is a systemic property of science taken as a whole, not a property of individual researchers or research. It is value-full not value-free.

Ackoff (1979) also directs his critique towards the fact that the developments in the first wave of systems thinking, have not been responsive to the changing needs of management brought about, to a large extent, by radical changes taking place in the environment in which they are practiced. This is closely related to Churchman's critique, as the first wave approaches fail to recognize changes mainly due to the narrow use of models to depict the system they are working with as models are always seen as "given". It is implicitly assumed that the environment of the system it is dealing with, the containing system, is deterministic in nature – hence is predictable in principle if not in practice – but that the system it is dealing with has choice, is purposeful (Ackoff, 1979), this itself is a contradiction. In Ulrich's (1987) terms, we cannot understand the meaning of the model if we do not understand the model environment.

In Ackoff's (1979) perspective, systemic operational research focuses on problems but real life managers are not confronted with problems that are independent of each other, but with dynamic situations that consist of complex systems of changing problems extracted from messes by analysis. Because messes are systems of problems, the sum of the optimal solutions to each component problem taken separately is not an optional solution to the mess. The behaviour of a mess depends more on how the solutions to its parts interact than on how they act independently of each other. But the unit in systemic operational research is a problem, not a mess. Taking this into account, there is no such thing as an optimal plan for, or design of, a purposeful system in a dynamic environment. This goes in hand with what is stated by Checkland (1985), as he describes the approaches of the first wave of systems thinking as talking in the language of problems and solutions directed towards eliminating

these problems.

The authors realized the need for a decision-making systems that could learn and adapt quickly and effectively in rapidly changing situations than the one there is for systems that produce optimal solutions that deteriorate with change, as well as a type of thinking that pays proper attention to the human component in socio-technical systems. That is how the second systems wave emerged, but this does not mean that first wave systems thinking was dismissed. Instead, Checkland (1985) viewed the hard methodologies as a special case of the Soft System thinking that was later developed.

2.3 Second Wave of Systems Thinking

According to Checkland (1985), the second wave of systems thinking was developed because the first wave, based on defining goals or objectives, did not work when applied to messy, ill-structured, real world problems. In this sense, the emphasis of the second wave is on how to cope with ill-structured problems or messes (Jackson, 1991a). It has at its core an interpretivist character rather than an objectivist one. Interpretive theorists adopt a more subjectivist approach to social science, they do not seek to study objective social facts and do not seek to search for regularities and causal relations in social reality. Instead the social world is seen as being the creative construction of human beings (Jackson, 1985). Hence the importance of probing the worldview (*Weltanschauung*) that individuals and groups employ in understanding and constructing the social world for people that matters.

These methodologies differ from the first wave methodologies in that they do not seek to reduce the messes to something that can be mathematically modelled. On the contrary, they seek to explore these messes by working with the different perspectives of multiple stakeholders, and has as a long term objective to encourage and institutionalize a process of continual learning among the participants of the social systems being addressed. In the second wave methodologies, systems are not physical objects but mental constructions of particular situations, for that reason multiple views of reality are admitted, values are included in the methodological process, and the immediate aim is to reach an accommodation about actions to be taken (Jackson, 1991a). These methodologies, accept that in social life there are multiple realities. For this reason, the models developed throughout the second wave do not pretend to be models of the world, only models that embody a particular way of viewing the world. It also differs from the first wave of system thinking in that the latter is focused on finding out about a problem in order to be able to solve it, while the former is

focused on taking action in the situation by carrying out explicit and organized systems thinking about a real-world situation. The work of C. West Churchman, Russell Ackoff and Peter Checkland is held to constitute the core of the second wave of systems thinking.

2.3.1 Developments of the Second Wave of Systems Thinking

From Churchman's view the analysis must respect different points of view concerning the goals to be attained by organizations and society. If a system is not serving its clients, the decision makers in the system must be persuaded to expose their worldviews to counter worldviews which represent the deadly opponents of their own worldview (Churchman, 1971). Thus, the systems approach begins when you first see the world through the eyes of the other. From Ackoff's view, there can be many different definitions of what the problem is in any particular situation. His approach is designed to bring about a consensus among the stakeholders concerned with a particular system so that action can be taken. This consensus, arrived at through open debate, is the only kind of 'objectivity' that can be obtained: "objectivity is the social product of the open interaction of a wide variety of individual subjectivities" (Ackoff, 1974). Consensus is sought through the process of interactive planning where stakeholders participate in designing an "idealized future" for the system with which they are concerned. Finally, from Checkland's view, methodology is not focused on any external "reality" but on people's perception of reality, on their mental processes rather than on the objects of those processes. The various worldviews operating in a system are expressed in Root Definitions of some of the systems which appear relevant to the problem situation, which explicitly name a number of features of the relevant system through the acronym CATWOE. By using Root Definitions, conceptual models are built. These are models of purposeful activity considered relevant to debate and argument about the problem situation. After developing these models, the debate about the situation is structured by comparing models (conceptual) with perceptions of the real world (rich pictures). The aim of the debate is to find some possible changes which meet two criteria: systemically desirable and culturally feasible in the particular situation in question. The debate which the second wave engineers through the comparison phase reveals the norms, the standards, the values extant in the problem situation and contributes to their changing (Jackson, 1985). This is a cyclical process as the definition of desirable and feasible changes gives a new problem situation on how to implement the desired changes (Checkland, 1985).

2.3.2 Criticism of the Second Wave of Systems Thinking

Although the second wave of systems thinking represented an advance from the initial developments of the first wave, there were two main concerns regarding it, which led to the further development of the third wave of systems thinking. First, that the interpretive theory underpinning it was inadequate for understanding and acting in social situations, where there were inequalities in power and economic relations. Second, the idea of complementarity in the form of the first wave being a special case of the second wave (as Ackoff stated) would lead to denaturing the first wave by the second wave tenets (Flood & Jackson, 1991a).

According to Mingers (1980) and Jackson (1982, 1987, 2000, 2003), one of the key features of the interpretive paradigm is that it is oriented towards regulation, concerned with understanding how order and cohesion are achieved and maintained. For this reason, the second wave systems methodologies can be seen more suitable for preserving the status quo than going beyond it. Furthermore, its subjectivism, is seen to prevent it coming to terms with structural features of social systems such as conflict and power (Mingers, 1984). As according to Jackson (1987, 2003) there is a willingness to take as given “compromises” and “accommodations” achieved within the confines of prevailing power structures.

Thomas and Locket (1979) argue that the second wave work is limited to the special case where conflict between people is not subject to the domination of any one interest. Even though it deals with purposeful systems by direct involvement in them, it has no explicit value commitments to resolve the difficulties over ends and means. Although it may have been these very difficulties that prompted their development from the first wave methodologies. The second wave authors do not see their methodologies as serving any particular social group or class. However, the application of systems ideas to social problems through action-research in a mutually acceptable ethical framework cannot be ideologically or politically neutral as it will reflect the interests of the dominant groups in society. Many of the uses of systems methodologies will be intended to promote changes to “improve the human condition”, and these can be based on a range of value commitments (Thomas & Lockett, 1979).

On Jackson’s (1985, 2000) perspective, the second wave of systems thinking lacks any theory capable of understanding the social structure which produce and support different worldviews or capable of indicating where the real levers of change in the social totality might be located, making it severely limited in terms of its effectiveness in bringing about change in social

systems. Second wave thinkers tend to work for change at the level of ideas which means that change can take place by changing the participants worldviews, but if worldviews are linked to other social facts (political and economic) in a constraining social totality, they may not be so easily changed. It is extremely difficult to change people's worldviews without first doing something about the political and economic structures which condition those worldviews (Flood & Jackson, 1991b). Some sort of understanding of the nature of the present totality, its laws of transformation and the existing blockages to change would be required before effective action could be taken.

On the other hand, having abandoned the formulation of systemic theories as the starting point of inquiry, it is not open for the second wave thinkers to employ the theoretical control as a check on their methods. This means that opposed to the developments that took place in the first wave of systems thinking, authors in the second wave detached from a theoretical justification for the development of their methodologies. Instead they rely on the subjective character of multiple worldviews of those involved. The participative criterion of validation depends upon all the stakeholders of a system being prepared to enter into a free and open discussion about changes to be made. Yet it is unrealistic to expect all stakeholders to be willing to enter into such debate, which means that full and effective participation cannot be established (Jackson, 1985). All participants need to have the same access to resources and information and equal say in discussion. The kind of unconstrained debate envisaged here cannot possibly take place mainly due to the effect of power in the possibility of having free unconstrained access to information, resources and being able to participate in an open discussion. If true participation is not taking place, then true validation cannot take place either. The second wave approaches do little to try to redress the effect of prevailing power structures (Jackson, 1985). This argument is expressed by Mingers (1980) in terms of Habermas critique, as he considers these approaches as maintaining and reinforcing the political status quo. This is related with the fact that according to the second wave of systems thinking, problems within society are not merely contingent but systematically created by the very structure of society then small scale problem-solving, cannot help but maintain that which is the very problem – society itself.

In general terms, the second wave authors are criticized for not being radical enough when they use their methodologies in social systems where inequalities in the distribution of resources and power exist. There must be an understanding that there are many social systems for which the second wave methodologies are inappropriate. Ulrich (1987, 1993)

makes a point regarding this subject, as he considers that at least the second wave of systems thinking has kept serving an instrumental rationality, whereas the inclusion of a practical rationality has been left aside. With this, he proposes a reformulation of systems practice that not only serves an instrumental rationality but incorporates the dialogical and the normative dimension of rational practice, and in doing so reformulating the concept of system towards a critical holism. This inappropriateness is the one that led to the development of Critical Systems thinking.

After the development of the second wave of systems thinking, conflict rose between the systems practitioners of both, the first and the second wave as both viewed their methodologies as the absolute truth, threatening to tear the systems community apart and leaving no room for methodological pluralism. In the realm of the critique directed towards the second wave approaches, emerged the third wave of systems thinking initially with two independent strands directed towards power and pluralism. Ulrich and Jackson, can be seen as the two major authors of each one of these strands, respectively. Both of these authors sought to subsume the work of the other until Midgley managed to create a synergy of both of their works in his work on Systemic Intervention in 2000.

2.4 Third Wave of Systems Thinking

The third wave of systems thinking, is an important and substantial development in the management and systems sciences. It shares the second wave critique of the first wave approaches but it is also able to reflect more fully upon the circumstances in which such approaches can properly be employed. Fundamentally, it locates major shortcomings in the second wave of systems thinking, particularly its failure to question its own theoretical underpinnings and to be reflective about the social context in which it is employed (Flood & Jackson, 1991a, 1991b; Mingers, 1980; Ulrich, 1993, 2003; Jackson, 2000, 2003).

As stated in the previous section, the third wave of systems thinking was developed in two strands focused on pluralism and power, respectively. These two strands will be described in what follows. Afterwards, I will describe an approach that can be seen as a third strand of the third wave, which managed to incorporate the work that took place in the two initial strands. It is important to clarify that although there are being named as a first and second strands, this is only done for narrative purposes as they took place in parallel and none is seen as being more relevant than the other.

2.4.1 The first strand of the third wave of systems thinking

This strand is characterised by being based on three themes or principles previously quoted by Flood & Jackson (1991a) as commitments. In what follows, we will be referring to these commitments as themes or principles. Throughout the years several authors have addressed these themes in different ways. According to Flood & Jackson (1991b), the development of the critical systems endeavour possesses three interrelated intentions, namely: complementarism, emancipation, and critical reflection. Midgley (1996a) and Schechter (1991) express these themes as critical awareness, emancipation and methodological pluralism, whereas Jackson (1991a, 1991b) initially expressed them in terms of five principles; critical awareness, social awareness, human emancipation, the complementary and informed use of systems methodologies in practice, and the complementary and informed development of all varieties of systems approaches. Midgley et al (1998) incorporating the previous work of these authors, defines these themes as critical awareness, improvement and methodological pluralism.

Flood & Jackson's view of pluralism, namely complementarism, is committed to reveal and critique the theoretical (ontological and epistemological) and methodological bases of systems approaches, and to reflect on the problem situation in which approaches can be properly employed and to critique their actual use (Flood & Jackson, 1991c). This position would respect the different strengths of the various trends in management science, encouraging theoretical development and suggesting ways in which they can be appropriately fitted to the variety of management problems that arise (Jackson, 1999). It is founded on both methodological and theoretical pluralism, although they are described as having different bases. In terms of Flood & Jackson (1991b), theoretical pluralism involves commensurability between perspectives, while methodological pluralism is based on the incommensurability of approaches. The term commensurability can be defined as being measured by a common standard; in other words, the reconciliation and integration of theories. In this context integration means that theories are denatured, being reduced in meaning to the fundamental tenets set by the measurement standard. On the other hand, the term incommensurability means that there is no common measurement standard, implying that there is no possible way of comparing theories (Flood & Romm, 1996b). Flood & Jackson's view of methodological pluralism has an instrumental and an epistemological underpinnings. The first one is related with the appropriate use of methodologies, while the second one is related with the use of Habermas' Knowledge constitutive interests' theory (Midgley, 1996a). Complementarism

makes explicit use of a metatheory to identify the strengths and weaknesses of different methodologies, and the methodologies as thereby viewed as complementary. Is in this sense that Jackson & Keys (1984) and Jackson (1991a) developed the System of Systems methodology, a pluralist metatheory to classify system methodologies according to the assumptions they make about social reality. It aligns the methodologies with the contexts that they should be capable of dealing with. This metatheory will be described in more detail later in this chapter.

According to Gregory (1996a), the complementarist strategy presented by Flood & Jackson (1991), cannot cope with dissensus about the nature and roles of the different systems methodologies, forcing a unitary view. This is explained by using a force field, where a more powerful position will “suck-in” the less powerful ones, subsuming them to their view. This means that the map of methodologies, and paradigms presented in the System of Systems Methodology has the potential to become stultifying, to close down debate with and learning from other perspectives, to fail to recognise insight about its own social domain, and to dominate interactions with and between other perspectives (Gregory , 1996). Midgley (1996a) states that a problem with Flood & Jackson understanding of methodological pluralism, is that it claims to be meta-paradigmatic as it coordinates the use of other systems paradigms. However, this cannot be the case given that the authors make assumptions about human knowledge drawn from the Knowledge-constitutive interests theory developed by Habermas. On the other hand, in Schechter (1991) terms, pluralism is committed to develop a broad framework which facilitates productive interaction between the advocates of different approaches, as all systems approaches have a contribution to make. Whereas Midgley (1996a) defines methodological pluralism as the use of a variety of research methods in a theoretical coherent manner and being aware of their weaknesses and strengths in dealing with corresponding variety of issues.

On the other hand, emancipation in terms of Flood & Jackson’s (1991a) view seeks to develop systems thinking and practice and to formulate new methodologies to tackle problem situations where the operation of power prevents the proper use of the newer soft system approaches. Jackson (1991a, 1991b), defines emancipation in terms of human emancipation, where the main goal of CST is to achieve for all individuals the maximum development of their potential. In Schechter (1991) terms, emancipation is a commitment to human beings and their potential for full development via free and equal participation in community with

others. It is also a commitment to recognizing the barrier to human emancipation – unequal power relations and the conceptual traps which perpetuate them – and incorporating this understanding into systems thinking. This goes in hand with the definition given by Midgley (1996a) where he states emancipation as a type of improvement that is defined in time and space and that takes into account power relations. In Midgley's (1996a) opinion, Jackson's interpretation of human emancipation will be directed towards the promotion of human well being separated from the consideration of the environment. The commitment to this type of emancipation, does nothing to encourage people to challenge the prioritization of human boundary in system practice, as this type of emancipation can only be achieved by addressing the three human interests proposed by Habermas.

Finally, critical awareness in terms of Flood & Jackson (1991) work, intends to reflect upon the relationship between different organizational and societal interests and the dominance of different systems theories and methodologies. It was initially divided in three types directed towards different things, the first one was directed towards the critical thinking towards methodology, the second towards the critical use of methodology and third one towards understanding power relations and improvement through ethical critique. Midgley (1996a) questions how the authors choose between the different critical awareness perspectives and how is the commitment to critical awareness enacted in situations where coercion is not identified. The lack of methodological aspects for critical awareness could have important implications. There could be a tendency to take the commissioning agenda as given and there could also be a failure to take sufficient account of the effects of that agenda on the organization's wider environment. He then proposes to define the commitment to critical awareness in terms of the ethical critique of boundaries, and to continue to conduct research to enhance both the theory and practice of boundary critique. Making up-front boundary judgments cannot be avoided in any research situation. Failure to realize the full implications of this will inevitably result in some of the most important boundary judgments – those that determine who the researcher will talk to and how the initial remit of the work will be defined – being made in an uncritical manner (Midgley, 1996a). Finally, Schechter (1991) describes the commitment to critique as a commitment to questioning the methods, practice, theory, normative content and limits of rationality of all-schools of thought. It requires a never ending attempt to uncover hidden assumptions and conceptual traps. While Midgley (1996a) takes critical awareness as being concerned with examining and re-examining assumptions that are usually taken for granted, as well as examining the conditions that precede them.

Taking into account this work, Midgley et al (1996a, 1998: pp.1) presents a concrete definition of each one these themes, as follows:

- *Critical awareness* - examining and re-examining taken-for-granted assumptions, along with the conditions that give rise to them;
- *Improvement* - defined temporarily and locally, but in a widely informed manner, taking issues of power (which may affect the definition) into account; and
- *Methodological pluralism* using a variety of research methods in a theoretically coherent manner, becoming aware of their strengths and weaknesses, to address a corresponding variety of issues.

It is important to recognize that when talking about pluralism, we can identify three types of pluralism; methodological pluralism, theoretical pluralism and pluralism regarding the nature of the relationships between the participants of a problem situation. Methodological pluralism can be understood as the use of different methodologies, methods and/or techniques from different paradigms in combination (Jackson, 1999; Midgley & Shen, 2007). Theoretical pluralism refers to drawing upon more than one theoretical “lens” to inform practice as implicit within different theories are contrasting themes, narratives and metaphors, which (when made explicit) can cast new light on a problematic situation (Midgley, 2011). While the third form of pluralism noted earlier refers to a situation in which participants have divergent values and beliefs and, to some extent, differing interests and objectives, but a genuine accommodation or compromise can be reached upon which all agree (Jackson, 1991b).

Taking into account the definitions given above about each of the commitments, we will now explore the most important developments that have taken place in the first strand of the third wave of systems thinking.

2.4.1.1 System of Systems methodologies

System of Systems Methodologies (SoSM) is one of several attempts to address pluralism through Critical Systems thinking. It was developed by Jackson & Keys (1984) to provide guidelines pointing towards the strengths and weaknesses of different systems approaches and suggesting which type of situation favours the use of one or another. It is a pluralist metatheory to classify system methodologies according to the assumptions they make about social reality. SoSM aligns the methodologies with the contexts that they should be capable of dealing with (Flood & Jackson, 1991a; 1991b). The key idea of SoSM was that it is

possible to construct an “ideal-type” grid of problem contexts that can be used to classify systems methodologies according to their assumptions about problem situations (Jackson, 1999). The SoSM offered a way forward from the prevailing management science in crisis debate, which saw different problem-solving approaches in competition, because it presented the different methodologies as being appropriate for different types of problem context. It established pluralism as a central tenet of critical systems thinking (Jackson, 2001).

Systems approaches are based upon different metaphorical understandings (machine metaphor, organic metaphor, neuro-cybernetic metaphor, cultural metaphor and political metaphor), and different views of reality and this conditions the way each advises seeking to intervene in and change organizations (Flood & Jackson, 1991c). This methodology, groups’ problem contexts according to two dimensions; systems and participants. A problem context is defined to include the individual or group of individuals who are the would-be problem solvers, the system(s) within which the problem lies and the set of relevant decision makers. This set contains all of the elements which can make decisions which may affect the behaviour of the system(s) (Jackson & Keys, 1984). The system dimension refers to relatively complexity in terms of the system or systems that make up the problem situation, and within which other difficult pluralistic or coercive issues of concern may be located. The participants dimension refers to the relationship between the individuals or parties who stand to gain from a systemic intervention.

Taking into account the systems dimensions, Jackson & Keys (1984) suggest a continuum of system types between simple and complex systems according to the characteristics that the system is showing. Simple systems are those that are constituted by a small number of elements, with few highly organized interactions between them, predefined attributes of the elements, that does not evolve through time and is largely closed to the environment. On the other hand, complex systems are defined as having a large number of elements, with many loosely organized interactions among them, with a probabilistic behaviour, that is continually evolving and that is largely open to the environment.

On the other hand, the participants are classified taking into account the type of relationships among them. There are three types of possible relationships; unitary, pluralistic and coercive. A unitary relationship means that participants share common interests, their values and beliefs are highly compatible and they largely agree upon ends, means and objectives. A pluralistic relationships means that although there might be different objectives, participants

are able to reach an accommodation between them. Finally, the coercive type of relationship is one in which participants do not share common interests, their values and beliefs are likely to conflict, they do not agree upon ends and means and genuine compromise is not possible, some coerce others to accept decisions and there is not an agreement over objectives.

Taking into account the previously described dimensions, Jackson & Keys (1984) developed a six celled matrix comprising 6 “ideal type” problem categories; simple-unitary, complex-unitary, simple-pluralist, complex-pluralist, simple-coercive and complex-coercive. Each of these problem contexts differs in a meaningful way from the others. The charting of these six ideal type problem context implies the need for six types of “problem solving” methodology.

2.4.1.2 Total Systems Intervention

Total Systems Intervention (TSI) was originally conceived to represent an approach to planning, designing “problem solving” and evaluation. It used systems thinking metaphors that are linked through the SoSM to encourage creative thinking about organizations and issues managers have to confront (Flood & Jackson, 1991a; 1991b). However, since it was conceived its authors have developed it in different directions. In the following section we will describe not only the original conception of TSI but also how it was updated in different ways in further work of Flood (1995, 1996) and Jackson (1999, 2001, 2003).

TSI sought to put in practice the principles of the third wave of systems thinking. TSI takes pluralism as complementarity where different methodologies express different rationalities from different alternative theoretical positions, these positions should be respected and methodologies and their appropriate theoretical underpinnings should be developed in partnership. In terms of TSI, it addresses pluralism through the use of metaphors and of SoSM at the core of its different stages. Social awareness (critical awareness) recognizes that there are organizational and societal pressures which have led to certain methodologies being popular for guiding interventions at particular times. This is embedded in the practice of TSI as it makes its users to be aware of the social consequences of using particular methodologies. Finally regarding emancipation, TSI has an emancipatory character in that it seeks to achieve for all individuals the maximum development of their potential (Flood & Jackson, 1991a, 1991b).

It regards problem situations as messes that cannot be understood and analysed on the basis of only one perspective, for this reason it advocates viewing them from a variety of perspectives perhaps as encapsulated in different metaphors. Once agreement is reached

among the facilitators and participants about the major issues and problems they are confronting, an appropriate choice needs to be made of systems methodology, or set of systems methodologies, for managing the mess and tackling the issues and problems (Jackson, 2003).

TSI used a range of metaphors to encourage creative thinking about organizations and the difficult issues they have to face, this metaphors were articulated through the SoSM framework to various systems approaches. TSI was based not only on the CST commitments but also on seven principles embedded in three phases; creativity, choice and implementation. It is important to clarify that these phases do not take place in a single linear way but they can take place several times and the decision makers can go back and forth through them, as much as he/she needs to.

The task during the creativity phase was to use systems metaphors as organizing structures to help managers think creatively about their enterprises. The aim was to take the broadest possible critical look at the problem situation but gradually to focus down on those aspects most crucial to the organization and that point in its history (Jackson, 2001). In order to do so, the tools provided by TSI were “systems metaphors”, different metaphors focus attention on different aspects of an organization’s functioning such as human and political aspects, functioning or organizational structure. The outcome of this phase was a dominant metaphor which highlighted the main interests and concerns and could become the basis for a choice of an appropriate intervention methodology (Flood & Jackson , 1991a).

The second phase was choice, the task during this stage was to choose an appropriate systems-based intervention methodology to suit particular characteristics of an organization’s situation as revealed by the examination conducted in the creativity phase. The tools for the development of this phase, were the guidelines of the SoSM and derived from that, knowledge of the underlying metaphors employed by systems methodologies. The most probable outcome of this stage, was a dominant methodology chosen to be tempered in use by the imperatives highlighted by dependent methodologies.

The last phase was the implementation stage, the task during this stage was to employ a particular systems methodology to translate the dominant vision of the organization, its structure and the general orientation adopted to concerns and problems, into specific proposals for change. In order to do so, TSI provided specific systems methodologies used according to the logic of TSI. The outcome of this stage was coordinated change brought

about in those aspects of the organization vital for its effective and efficient functioning (Flood & Jackson , 1991a).

According to Jackson (1999), the contribution of TSI is to postulate a meta-methodology for using methodologies adhering to different paradigms in the same intervention on the same problem situation. TSI was successful in providing guidelines for the use of critical systems ideas in practice, in that it employed critique of different systems approaches, respected the possibility of “coercive” contexts and was based upon a sophisticated form of pluralism in which methodologies adhering to different paradigms were to be used in the same intervention on the same problem situation (Jackson, 2001).

However, after its initial conception, Flood, and Jackson along with other authors recognized several shortcomings in the methodology. Jackson (1999) regards as a weakness of TSI that it grounds its pluralism or “complementarism”, uncritically on Habermas’ early theory of human interests, this goes hand in hand with that express by Midgley (1996a). Another gap recognized by Jackson (1999) is highlighted by the fact that although TSI demand multi-methodological competences it does not detail how the relevant competences can be obtained. Midgley (1992a), Dutt (1994) and Mingers & Brocklesby (1996) noted that TSI emphasises the use of whole methodologies, which means that there is an unnecessary lack of flexibility which needs addressing. Jackson (2001) states that this flexibility could be gained by using methods, models and techniques (perhaps normally associated with different methodologies) in combination.

In response to the critics, Flood decided to reframe the initial conception of TSI by proposing the Local Systems Intervention (LSI) which constituted a reconstruction of TSI taking into account the critique directed towards its initial version. LSI is an approach that encourages diversity through local decision making about the relevance of designs (How), considerateness of decisions (What) and astuteness of judgment made about designs and decisions (Why) (Flood, 1996). The aim of LSI is to put in place a system of thought for action by exploring the three discourses (How, What and Why) that allow practitioners to recognize and account for the response that they might make when addressing problematic situations.

LSI can be studied through a set of principles, philosophy and process. Philosophy informs how the world looks when a LSI perspective is assumed. LSI comprises two branches of philosophy, one focused towards when LSI is directed towards an organization and another

focused on when LSI is applied towards organizational problem solving. When applied to organizations, LSI provides a systemic view of them. Organizations comprise parts that are continually interacting, they form a system (horizontal) that is a subsystem of a larger system and has subsystems itself (vertical). An organization is therefore both vertically and horizontally integrated. At each systemic level the phenomenon emergence occurs, meaning that the result of the whole activity is more than the sum of the results of activities which any level comprises (Flood, 1995). This image of organizations provides a framework on which an ideal whole system view can be constructed, with five stages of construction. When applied to organizational problem solving, TSI extends its systemic interpretation of organizations. Problem solving means managing sets of interacting issues, it is a particular type of human activity that is, by definition, a part of the interacting activities. Taking this into account, organization is understood in terms of interacting issues whereas problem solving, or intervention, as being part of a continuous process of managing these issues through the whole organization, taking into account wider concerns.

LSI principles propose kinds of action we should take, it is possible and only possible to evaluation a LSI process using its own principles. LSI includes four main principles: being systemic, achieving meaningful participation, being reflective, and striving for human freedom (Flood, 1995). The systemic principle means that the system under study is a part of a greater whole and that it comprises interacting parts and in the particular case of LSI, it operates at a system, subsystem and supra-system levels. The principle of meaningful participation means that the perception of all people involved and affected must be drawn into the picture. The principle of reflection includes two reflective needs with LSI: to reflect upon the relationship between different organizational interests in order to gain a whole system understanding and to reflect upon the dominance of favoured approaches to intervention to ensure that all issues are managed. The achievement of both needs promotes effective management. The fourth principle introduces LSI's ideology, it is a principle that stresses achieving freedom for all. This principle can only be achieved by following the preceding three.

Finally, the process sets out how to implement the four principles, it has three phases: creativity, choice and implementation. Each of these phases, supports and is supported by each other phase, through a continuous process with no predetermined start or finish points. These phases are the same as the ones expressed in the initial development of TSI which is why they will not be detailed described. Instead, the description will be focused on the

differences among the stages of both methodologies. LSI stages can take place in a clockwise or a counter clockwise mode. The clockwise mode works as the original TSI mode where the outcome of each phase passed to the next phase to accomplish a specific task, whereas the counter clockwise mode works as process of critical reflection that questions the outcome of the previous stage (Flood, 1995).

Regarding the critique towards TSI that led to the development of LSI, Flood (1995) addresses this referring to the design of each of the stages of the new methodology. First, Flood (1995) refers to the critique directed towards the creativity stage, here one of the biggest concerns regarding TSI was the limited number of metaphors available to conduct creative thinking. As a response to this critique Flood introduced further methods for creative thinking, such as brain storming and lateral thinking and divided the creative process into decontextualizing and contextualizing. Regarding the choice phase, critique was directed primarily towards the difficulties of using SoSM, which led to the redevelopment of SoSM in a “complementarist framework”.

Later on, Jackson (1999, 2001, 2003) proposes the need for a reformulation rather than the abandonment of TSI, he presents a modified approach called “critical systems practice”. In critical systems practice, a meta-methodology is required which encourages and protects paradigm diversity and handles the relationships between the methodologies, based on alternative paradigms, in order to address the complexity and heterogeneity of problem situations at all stages of an intervention. The meta-methodology accepts that paradigms are based upon incompatible philosophical assumptions and that they cannot, therefore, be integrated without something being lost. It seeks to manage the paradigms not by aspiring to meta-paradigmatic status and allocating them to their respective tasks, but by mediating between the paradigms (Jackson, 2003). However, in terms of the commitment to the pluralism, critical awareness and emancipation themes in still at the core of the methodology, as well as the commitment to the initial principles and possessing the three general phases of creativity, choice and implementation.

As could be seen here, although TSI was developed by Flood & Jackson (1991), years later both authors abandoned their initial conception of TSI and presented further developments in the form of Local Systems Intervention and Critical Systems Practice which were developed by Flood and Jackson respectively.

2.4.2 Second strand of the third wave of systems thinking

The second strand of the third wave of systems thinking has developed by Werner Ulrich in the 1980s inspiring in the work of Churchman (1970) regarding the conception of the system and is represented in the development of a social theory and methodology called Critical Systems Heuristics.

2.4.2.1 Critical Systems Heuristics

CSH lies on two traditions of thought; systems thinking and practical philosophy. As stated previously, Ulrich inspired in the work of C. West Churchman in the systems thinking subject. Throughout his work, Churchman recognizes that boundaries are personal constructs that define the limits of the knowledge that is to be taken as pertinent in an analysis. As different people may use different boundary judgments defining what is to be included in analyses, and whose views have credibility, what looks like an improvement through one pair of eyes may look like the very opposite through another (Churchman, 1968, 1971). This sets the question of the view (lens) in which one might see the world differently. In CSH such a lens is referred to as a reference system – a conceptual device circumscribed by the 12 boundary categories which in turn are defined by the boundary questions for which they stand.

While Churchman preferred to view systems as real-world entities, he nevertheless provided the initial foundation categories adapted later for delineating a CSH reference system. He first identified nine ‘necessary conditions’ for conceiving anything as a system, and later extended these to 12 ‘planning categories’ (Ulrich, 1983). The three additional conditions are ‘systems philosophers’, ‘enemies of the systems approach’, and ‘significance’; Churchman understands them to raise issues related to the significance of a systems perspective as distinguished from the partial (because non-holist) perspectives of the ‘enemies’. In strict CSH terms, they are critical ‘ideas’ for meta-level reflection about the meaning of a systems approach rather than ‘categories’ for mapping any specific system. Despite this difference of understanding, Churchman’s theme of the ‘enemies’ pointed the way to CSH’s notion of boundary critique, in an effort to give a precise methodological meaning to his credo ‘know (and love) thy enemy’. It also engendered an important heuristic device of CSH, the idea of maintaining tensions between contrasting perspectives for critical purposes; “we have to maintain the contradiction or else we allow ourselves to be overwhelmed by the consistent” (Ulrich & Reynolds, 2010).

The second main tradition picked up in CSH is the tradition of practical philosophy. This comprises two largely independent strands of philosophical thought. On the one hand, there is the tradition of American philosophical pragmatism and on the second is the critical social theory. Both strands of practical philosophy are to an important degree rooted in Immanuel Kant's (1787) critical philosophy, from which CSH derives many of its central concepts.

The pragmatic perspective of CSH means that it is oriented towards practical rather than theoretical ends; accordingly, CSH employs an action-theoretical framework, that is, it looks at situations from the point of view of an agent rather than an observer. Its critical perspective means that CSH considers values – and value conflicts – as integral part of all claims to rational practice; it relies on a discourse-theoretical (or 'discursive') framework to assist users in dealing openly and critically with the value implications of boundary judgements.

As stated earlier, grounded on Churchman's statements about boundary judgments, Ulrich discusses how the exploration and setting of boundaries can be undertaken through dialogue between stakeholders as a way of rationally justify the boundaries they use. In order to do so, Ulrich relies on the work of Habermas (1976) concerning the nature of rationality. Habermas regards rationality as being dialogical, with language as a tool which allows the participants of a situation, open and free questioning between them. However, Habermas acknowledges that dialogue may be distorted through the effects of power which can present itself in a direct or indirect manner when one participant coerces another or when participants make unquestioned assumptions about the absolute necessity of, or the inevitable future existence of particular social systems. To overcome these effects of power, we need to establish what Habermas calls an 'ideal speech situation': a situation where any assumption can be subject to critique and all viewpoints can be heard (Midgley , et al., 2007). The ideal speech situation excludes systematic distortion of communication. More precisely, the structure of communication produces no constraints if and only if there is a symmetrical distribution of the chances of all participants in the discourse to select and perform speech acts (Habermas 1973, 1970c, 1971c in Ulrich 1983). Speech acts are elementary units of linguistic communication, e.g. making a statement, asking questions, giving commands, etc. In every speech act, we communicate with somebody about something and thereby also express our own subjectivity. Habermas finds the normative fundament of discourse in four unavoidable validity claims; comprehensibility, truth, truthfulness, and rightness, each of which comes into play in every instance of discursive rationality, though they cannot all take place simultaneously.

Introduction

Ulrich developed the Critical Systems Heuristics (CSH) as a way to respond to the necessity of reframing the concepts of citizenship and social planning in modern society. Ulrich (1996a, 1996b, 1998) proposes to develop a citizenship status that was not only focused on civil rights but also on the development of civil competences, as in this way the role of citizens can move towards “active citizenship”, this is an idea in which ordinary people have an effective and equal chance of participating in the making of public opinions and political decisions. Ulrich argues that although this “active citizenship” might be an ideal, in order to achieve it, it is necessary to translate systems thinking into a language that ordinary people can understand and are willing to use, and which in turn will empower them in a meaningful way to be able to control their lives. In order to so, he sought to find a source of competence in citizenship that would be independent of any specific expertise, but that aside of giving people a voice, would give them a competent voice as well.

This conception of citizenship implies a turn in the concept of social planning as well. Social planning is traditionally practiced as a form of purpose rational action, that uses scientifically informed means for accomplishing a given end (Ulrich , 1996b). In contrast, CSH understands it as the art of promoting improvement. The problem with the notion of improvement, is that there is no objective way of determining if something constitutes an improvement or not, as the concept relies on the knowledge and values of the people that are being involved in a situation, which means that what might constitute an improvement for someone might not be seen as such by another person. These are what Ulrich (1983, 1987, 1998, 2005a) calls boundary judgments, they determine which facts (empirical observations) and values (value considerations) count as relevant and which others are left out or considered less important, they define the boundaries of the reference system that is constitutive of the meaning of a proposition and for which it is valid. These boundary judgments, are constitutive of the reference systems to which refer all our claims to knowledge or rationality, in professional practice as well as in everyday life (not only in improvement). The way to tackle the subjective character (in terms of partiality) of improvement and in general terms everyday life situations, was to address boundary judgments through a critical employment of the systems idea, this is through boundary critique.

The combination of systems thinking and of practical philosophy has enabled CSH to

influence critical systems thinking. According to Ulrich and Reynolds (2010), the point of departure for a critical systems approach lies in the notion that all uses of approaches, methodologies, and methods whether described as systems or something else, are partial, in a dual sense. First, by representing only a section rather than the whole of the total universe of possibly relevant considerations, and second, by serving some parties better than others as their uses are driven by value judgments. The implication of this viewpoint is that using a systems approach requires a systematic consideration of what a system map or design may leave out (always knowing that the analysis cannot be comprehensive) and to always examine them from multiple perspectives. In that sense, the development of CSH is a critical systems approach developed to embrace this dual sense of partiality.

Although Ulrich recognizes CSH as being part and contribution to CST, his understanding of CST distinguishes from the traditional description of CST in terms of “commitments or themes”. As he states that anyone can claim such things as critical awareness and dedication to emancipation; the point is to translate them methodologically (Ulrich W. , 1996a).

CSH and Boundary Critique

Boundary Critique is the methodological core idea of CSH (Ulrich, 1983). It is defined as a systematic – reflective and discursive – effort of handling boundary judgements critically, whereby ‘critically’ means both ‘self-critically’ questioning one’s own claims and ‘thinking for oneself’ before adopting the claims of others (Ulrich & Reynolds, 2010).

Boundary critique involves first of all a process of unfolding, that is, making ourselves and others aware of the boundary judgments assumed with respect to 12 kinds of boundary judgments. This is a specific tool for uncovering the inevitable selectivity of all our systems maps and designs; that is, it serves a critical purpose against all holistic pretensions (Ulrich , 2004). In general terms, this process takes place as follows. First, the need to examine different boundary judgments is addressed through the use of twelve boundary questions in the “ought” mode, which is a normative ideal mode, and the “is” mode, a descriptive and more realistic mode. Second, judgments can be contrasted among the stakeholders groups associated with the sources of influence (motivation, power, knowledge and legitimation), these categories will be explained later. Third, judgments can be contrasted between those stakeholders “involved” in the systems design and those “affected” by its consequences but who are not involved. Fourth, it is possible to review the entire set of boundary judgments associated with any reference system in the light of another set of boundary judgments

belonging to a different reference system. The development of an unfolding process at a practical level implies a series of key concepts, namely: the systematic alteration of boundary judgments, the alternative planning roles, an ethical boundary critique and the possible ideal of a convergence of the role-categories (Ulrich , 1996b).

The systematic alteration of boundary judgments emerges as the process of unfolding consists in using the critically-heuristic categories or boundary questions for a systematic, iterative expansion and narrowing-down of the considered context. This helps to identify boundary judgments and to surface and debate their value content, by tracing and contrasting the practical implications of alternative boundary judgments. This process is basic to self-reflection, mutual understanding and debate on the meaning of improvement, as it helps reveal the partiality of any design of improvement. The alternative planning roles of the unfolding process contribute to the achievement of a systematic expansion and narrowing-down of boundary judgments as it conceptualizes the role of planners in different viewpoints. This process is also meant to help overcome a “pluralistic” position that would ultimately imply a bottomless relativism of values, as it requires a step-by-step inclusion of more stakeholders beyond that of the directly involved people. Finally, the essence of the unfolding process lead to the convergence of role-categories, in the sense that different stakeholders such as the decision-maker, the witness and the planner can be alternatively by citizen planners, as the ultimate goal of CSH is providing ordinary citizens with tools to make them competent at the same level of trained planers and decision makers, in subjects where knowledge does not represent a real advantage.

In addition to the process of unfolding, systematic boundary critique involves a second effort, the systematic questioning of boundary judgements with respect to their adequacy in terms of relevance, justification, and ethical defendability. Whereas the aim of ‘unfolding’ consists in uncovering the selectivity of the reference systems at work in our claims, the aim of ‘questioning’ consists in exploring and, if necessary, challenging their resulting partiality. To this end, boundary questioning requires that we thoroughly analyse actual and possible consequences and ask what they may mean for all the parties concerned; and furthermore, that we examine what options may be available that might provide a better basis for mutual understanding and cooperative action towards ‘improving’ the situation.

According to Ulrich (1983), the quest for systemic thinking cannot alter the fact that all our claims remain partial, in the double sense of being selective with respect to relevant facts and

norms and of benefiting some parties more than others. This is what boundary critique is all about; it aims at disclosing this inevitable partiality (Ulrich , 1996b). The main vehicle driving this critical process is the critical employment of boundary judgments. The idea, is that boundary judgments offer themselves for three kinds of critical employment

A systematic process of boundary critique is to achieve three basic requirements, tasks (Ulrich, 2005a; 2005b), critical employment (Ulrich , 1998) or endeavors (Ulrich & Reynolds, 2010):

1. Reflective practice through critically-heuristic self-reflection. This is, to identify the sources of selectivity by surfacing the underpinning boundary judgments, implying an analytical focus on understanding situations, by revealing to ourselves and to others the boundary judgments at work and allowing everyone to understand their implications.
2. A dialogical search for mutual understanding and possible consensus through critically heuristic deliberation. Refers to the necessity of questioning boundary judgments with respect to their practical and ethical implication and to surface options, allowing a reflexive practice and a practical focus of improving a situation through discussions with all the concerned stakeholders.
3. Controversial debate through the polemical employment of boundary judgments. This is, to challenge unqualified claims to knowledge or rationality by compelling argumentation, through the emancipatory use of boundary critique. This “polemical” use of boundary judgments aims to make visible the operation of power, deception, and dogmatism of other non-argumentative means behind rationality claims. It accomplishes this purpose by creating a situation in which a party's reliance on non-argumentative means to support boundary judgements becomes apparent.

All three types of boundary critique can help people deal with the empirical ('facts') and normative ('values') selectivity of propositions in a competent and critical way, by enabling them to make transparent to themselves and to others how both 'facts' and 'values' depend on the choice of systems boundaries. Their optional character, i.e., the availability of alternative ways to bound the system of concern, along with the unavailability of any 'objective' justification for the chosen boundaries of concern, should become clear and the normative presuppositions and conceivable consequences of all options should be visible (Ulrich , 1998).

Ulrich offers a conceptual framework to explain the entire process relating these three tasks (Figure 1). This framework is called the “eternal triangle” of boundary judgments, observations and evaluations. The eternal triangle in Figure 1, shows that whenever we change one of the angles of the triangle, the other two angles will change as well.

Figure 1. The interdependence of boundary judgments, observations, and evaluations.

As stated previously, the concept of boundary judgments says that the meaning of a proposition depends on how we bound the relevant reference system, that is, the context that matters when it comes to assessing the merits and defects of the proposition. The assessment of merits and defects of a proposition depends on both observations of fact and judgments of value. What facts we observe and equally what values we judge appropriate, in turn depends on our reference system (Ulrich , 2000). When we change our boundary judgments the relevant facts and values change too, conversely the appearance of new facts can make us change our boundary judgments, leading to a reframing of relevant values.

Ulrich then developed a set of twelve basic boundary problems used in through twelve boundary questions to support the achievement of one of the boundary critique tasks previously stated as these questions can be used to achieve the analytical focus, the practical focus or the emancipatory focus of each of the tasks, according to the way they are used.

Boundary Questions

Prior to the development of the boundary questions, Ulrich developed twelve basic critically heuristic boundary problems. These problems represent twelve categories of “relative a priori” judgments methodologically grounded in a reconstruction of Kantian a priori science. Ulrich (1983, 1987, 1998), recognizes an a priori character because they come temporarily prior to analysing an issue, whereas he also recognizes a relativistic character in that they do

not come prior to every form of possible experience but only to the experience and evaluation of a particular issue.

Figure 2. Critical Heuristic Categories.

The categories of CSH shown in Figure 2, are derived from sociological considerations addressing the social actors whose views and intentions are constitutive of the system of concern. These categories, focus primarily on people, rather than on material, space or time, although these types of boundary judgments are equally relevant. Ulrich (1996b) grounds this decision on the fact that it is people who possesses the sense of purposefulness, power, knowledge and sense of responsibility, they together determine what ought what counts as an improvement in social planning. Taking this into account, the categories were derived from a “dimensional analysis” of the system of concern (context of application) that is relevant for determining the meaning of improvement. This analysis evolved around the distinction of “those involved” and “those affected” but who are not involved (see Figure 2). Regarding the importance of the system of concern, Ulrich (1998) states that two types of environment and one type of context could be recognized as being relevant for CSH, these are: the relevant environment, which does not belong to the system but influences it, the irrelevant environment, which does not belong to the system and neither does it influence it, and the context of application which constitutes that part of the universe where which is influenced by the system.

Different types of actors were identified, taking into account the distinction between those involved and those affected, these are recognized as; those who benefit, those who have the

say, those who have the knowledge, and those who merely have to bear the side-effects and costs. Acknowledging the existence of these actors, Ulrich proceeds to identify the specific types of concerns they contribute to the determination of improvement. The last consideration was done regarding the crucial issues that can arise from the clash of concerns with those of other social actors, as the way in which clashes are handled determines the meaning of improvement, as well. To all these constituents, categories that are meant to stand for the justification break-offs that they imply were then assigned.

These twelve boundary categories, are then put in a set of questions. These questions are arranged in four groups of three questions each. All questions shall be asked in an “ought” mode and in an “is” mode, where the “ought” is usually asked first as it provides us with a standard for subsequently evaluating the “is”. This is, as both modes are closely related; only together can they unfold their critical power and drive the process of unfolding a plan’s normative and empirical content (Ulrich 2000, 2003). The four groups of questions are intended to address the following sources of human intentionality: sources of motivation, sources of power, sources of knowledge, and sources of legitimation (see Figure 2).

The sources of motivation concern with the sources of value orientation and purposefulness that ought/is to underpin the plan, or, better, should be built into it. The idea is that by understanding the plan’s sources of motivation, it will be easier to determine what ought/is to be the plan’s value assumptions and, even more importantly, its implications for practical life (Ulrich, 1996a, 199b). The second group of question concerns with the sources of control, the sources of decision power and the means that ought/is to be available to realise the plan. By understanding the sources of control, it will be easier to determine what ought/is to be the plan’s assumptions regarding the scope of the plan and, what it should leave out of consideration (ibid). The sources of knowledge, concern with the sources of expertise, information and know-how, experience and skills that ought/is to be considered as relevant. This consideration will make it easier to understand what ought/is to be the plan’s assumptions regarding relevant facts, i.e., what ought/is to count as 'knowledge' and, even more importantly, what facts need not be considered to constitute relevant 'knowledge' and hence can be safely ignored (ibid) . Finally, the sources of legitimation concern the questions of what ought/is to be the sources of justification in relation with those people who are or may be affected without being involved. The idea is that if we understand these sources of legitimation, it will be easier to determine what ought/is to be the plan's ethical assumptions and, even more importantly, what ethical conflicts should be dealt with.

Each one of these categories, includes three types of question or sub-categories. The first question or sub-category (see numbers 1, 4, 7, and 10 in Figure 2), of each category refers to a social role (rather than an individual person) who is or should be involved in defining the system of concern. The second question or sub-category (see numbers 2, 5, 8, and 11 in Figure 2), addresses role-specific concerns that are or should be included. And the third question or sub-category (see numbers 3, 6, 9, and 12 in Figure 2), are crucial for understanding the previous two boundary judgments.

The four categories and the three sub-categories (or three questions in each category) are interdependent as the answer to one group of questions helps determine the answers to the other groups. Together the four groups of boundary judgments provide a “rich picture” of a plan’s fundamental anatomy of purposefulness (Ulrich , 1996b).

CSH Critique

Regarding the development of CSH, Midgley (1992, 2000) argues that a systemic intervention is more than the simple inclusion or exclusion of stakeholders and issues, as different groups might ground on different ethics (values in action) towards the same issues, the differences between boundary judgments might lead to the appearance of conflict. He argues, that is more common than rare to find situations where particular groups and their concerning issues are being marginalized in way where they are not completely excluded nor included in the system of concern.

We will explore these ideas in the following sections.

2.4.2.2 Boundary critique, boundary judgments and marginalization

According to Ulrich, if rationality is dialogical, plans for improvement should, in principle, be normatively acceptable to all those participating in a given dialogue (Midgley , Munlo , & Brown , 1998). Taking this into account, Midgley (1992, 2000) sought to extend Ulrich’s work on boundary critique as he argues that the consideration of different stakeholders and issues related to them in a systemic intervention implied the existence of different boundary judgments grounded in different sets of values, differences that might lead to the appearance of conflict. More precisely, he wanted to answer the following question: “what happens when two groups have different and conflicting boundary judgments about the same situation and the conflict becomes entrenched?” (Cordoba & Midgley , 2008). Is in answering it that the concept of marginalization emerges.

Midgley (2000) argues that if one group makes a narrow boundary judgment and another makes a wider one, there will be a marginal area between the two boundaries. This marginal area will contain elements that are excluded by the group making the narrow boundary judgment, but are included in the wider analysis undertaken by the second group. He names these, the primary and the secondary boundary, where the narrower boundary is the primary one. The representation of these boundaries is shown in Figure 3.

Figure 3. Marginalization.

As an answer to the question regarding conflictual ethical boundary judgments, Midgley (2000) states that when two ethical boundary judgments come into conflict, the situation tends to be stabilized by the imposition of either a sacred or a profane status on marginal elements. It is important to clarify several elements of this statement. First, in order for something to be considered as being marginalized, the existence and employment of multiple (at least two) boundaries is required as this is not a one sided process. Second, the character of the terms “sacred” and “profane” do not take a religious connotation. Instead, Midgley (2000) clarifies that they refer to a special status given to marginalized elements, referring to them being valued or devalued respectively. Third, the stabilization of conflict does not mean that the conflict is solved. The Oxford Dictionary (2017), defines stabilization as the process of making something more secure, stable or something that is unlikely to be overturned. This means that in the particular case of the processes taking place in a boundary judgment conflictual situation, the consideration of a boundary is imposed over the other. Then the imposition of a different boundary implies the stabilization of the conflict in different ways as it establishes a profane or a sacred status over marginal elements. When marginal elements

are sacred, the secondary boundary and its associated ethic is focused upon and reinforced. Whereas, when marginal elements become profane, the primary boundary and its associated ethic is focused upon and reinforced as the main reference for decision making. People or issues relegated to the margins are disparaged, allowing the secondary boundary to be ignored (Midgley , 2000).

There is rarely a consensus about whether marginalised issues or people should be seen as sacred or profane, but one of these interpretations generally comes to dominate – especially when it is institutionalised in social ritual (Cordoba & Midgley, 2006, 2008). According to Midgley (2000), ritual is behaviour, in whatever context, it contains certain stereotypical elements that involve the symbolic expression of wider social concerns. An observation of the presence of ritual can tell us where sacredness and profanity might lie, and hence where ethical conflicts related to marginalization might be found.

As can be seen in Figure 4, the consideration of a primary and a secondary boundary imply the emergence of a set of ethics arising from within each of the boundaries, these sets of ethics are the ones that will come into conflict and are the one which will be express symbolically through rituals. The only way of solving the conflict is by the imposition of one of them, giving a profane or a sacred characters to the marginal elements. The whole process is symbolically expresses in ritual which, in turn, helps to support the total system (Midgley , 2000).

Figure 4. Margins, ethics, sacredness, profanity and ritual.

At this point, Midgley (2000) problematizes Ulrich (1983) idea of boundary judgments being normatively acceptable to all concerned citizens. Midgley (2000) argues that although it is possible to have a consensus over a group of stakeholders about what boundary they want to adopt, this might be the result of processes of marginalization and conflict that remain invisible. For this reason, he stresses the importance of realizing what might be marginalized when all stakeholders share the same commitment to a boundary judgment. Midgley (2000) suggests the idea of involving stakeholders beyond those concerned and affected, stakeholders that are committed to different boundary judgments and that might challenge the existent consensus raising awareness about the marginalization that was taking place underneath it.

By the problematization of Ulrich's idea, Midgley extends the boundary critique theory as he does not limit himself to only exploring how conflicting boundaries operate in a single group of stakeholders but he also addresses how this situation might take place when different stakeholder groups with different purposes but with some common concerns come together. The representation of this type of dynamic widely differs from the one of a single group of stakeholders, as now the relationships between this groups is seen as being overlapped.

Figure 5 shows the relationship between two stakeholder groups. In this case, there are two groups of stakeholders (actors) with conflicting ethics, but in this case there is no marginal area as there are not primary or secondary boundaries within each one of the groups. This does

not mean that profanity or sacredness cannot take place, they do, but now what becomes sacred or profane is the group and its concern.

Figure 5. Model of overlapping stakeholder concerns, with attributions of profanity.

In Midgley's (2000) terms, when one group sees the other as sacred, and there is no element of profanity in the picture, it is unlikely that anyone will view the conflict as a major problem because there is an incentive for the group which sees the other as sacred to strive for co-operation – even to the extent of moderating their own values and ethics to come into line with what are perceived as sacred expectations. Therefore, while acknowledging that it is possible for one stakeholder group to view another as entirely sacred, I will put this situation to one side and concentrate attention on what happens when stakeholder groups regard one another as profane. This is when participants are most likely to experience their situation as problematic.

It is important to note that there is not a predefined attribution of sacredness and profanity but it is possible to obtain a wide range of sacredness and profanity as stakeholders might initially see each other in a position but later change the way they see each other to a completely different position. In this terms, Yolles (2000) proposes power relations and the subjective relations of actors as important elements to understanding marginalization processes. Midgley (2000), stresses that power relations come into play when one set of stakeholders is able to pursue its purposes through action in a manner that constraints the actions of the other stakeholders. On the other hand, the subjective character of the stakeholders are important because the exercise of power will only become problematic if a

set of stakeholders sees it as such.

Midgley (2000) argues that boundary critique can support the practice of methodological pluralism by being explicit about values and power relations and requiring more penetrating explorations of problematic situations than most problem structuring methodologies ask for (including some other methodologies that are explicit about the need for methodological pluralism). He suggests that explorations informed by the theory of boundary critique almost always result in the design of multi-layered interventions, involving the creative mixing of methods, which is in marked contrast to some other pluralist methodologies where more superficial diagnostic strategies tend to result in simple choices between ‘off-the- shelf’ methods (Foote, et al., 2007).

2.4.3 The third strand of the third wave of systems thinking

In 2000, Midgley developed a methodology called Systemic Intervention. Systemic Intervention is defined as purposeful action by an agent to create change in relation to reflection upon boundaries (Midgley , 2000). This definitions rests upon the separate considerations of the concepts of intervention and systems. Midgley (2000, 2014) conceives intervention as purposeful action by an agent to create change. Whereas he considers the boundary concept as being at the core of systems thinking mainly due to the interconnectedness of everything in the Universe, meaning that the place where a boundary is located is crucial – a boundary judgement always cuts interconnections and needs to be justified (as claimed previously by Ulrich, 1983).

Midgley (2000; 2008) suggests that an adequate methodology for systemic intervention should be explicit about three things: to reflect critically upon, and make choices between boundaries, this is boundary critique; the need for agents to make choices between theories and methods to guide actions, this is theoretical and methodological pluralism; and being explicit about taking action for improvement. First, establishing a boundary implies the existence of a particular set of values and stakeholders, being critical upon this, is the only way to analyse the ethical consequences of the different possible actions.

Second, setting a particular boundary not only requires being critical upon it, but it will also determine which theories and methods can be used in each case. Choosing between boundaries and theories suggest which methods might be most appropriate, and conversely, choices between methods imply particular theoretical and boundary judgments (Midgley , 2008). Third, Midgley (2000), acknowledges the importance of the concepts of action and

improvement as both have a subjective character in terms that they cannot be defined in an absolute manner. The meaning of action can only be determined in local contexts (local does not mean small) as the use of different boundaries, theories and methods will give rise to different understandings of what it means for an agent to take action. Second, improvement has a local and a temporary character as what might seem to be an improvement to one agent might be seen in a different way by another. Furthermore, the temporary character of an improvement is linked with it being able to acquire a sustainable character, in the sense that it can be achieved with long lasting results without the appearance of undesired consequences.

These three components of a systemic intervention have an inseparable character (at least in a factual character). According to Midgley (2000), doing one always implies doing the other two as well. However, the focus of attention may shift from one to another aspect so that none remain implicit and escape critical analysis.

Figure 1. Three aspects of a methodology for systemic intervention. Taken from Systemic Intervention: philosophy, methodology and practice.

Figure 6 shows the relationship between the activities as represented by Midgley (2000). Here the term critique refers to boundary critique, the term judgement refers to methodological and theoretical pluralism and the term action refers to the actions that need to be taken for improvement. We have previously stated that these components have an inseparable character. However, a separable character emerges at an analytical level. This separation ensures a proper consideration of a minimum set of three “angles” on possible paths for intervention (Midgley , 2000).

The main added value of systemic intervention compared with earlier systems approaches is its synergy of boundary critique and methodological pluralism. If a complex issue is defined from only one limited perspective without reflecting on values and boundaries, and issues of marginalization are neglected, then the outcome could be the use of a systems approach that misses or even exacerbates significant social problems. The synergy of boundary critique and

methodological pluralism ensures that each aspect of systemic intervention corrects the potential weaknesses of the other (Midgley , 2015). In the developments of the previous strands of the third wave of systems thinking, the most relevant authors of each strand, namely Ulrich and Jackson tried to devalue in some sense the work of the other. However, the development of Systemic Intervention shows that there was never a need to do that, not only because both developments could coexist by being used with different scopes but also because they could be integrated in a methodology that enriched both works.

In the next chapter, I will describe how the program evaluation field developed.

Chapter 3 – Evaluation Theory

Evaluation is a process that takes place in almost every daily life activity of an individual. However, it was only formally established as a field of enquiry in the midst of the twentieth century. This chapter seeks to explore and describe how the modern view of evaluation field was established and how it has been evolving for the past few decades by primarily reviewing the developments related to program evaluation. In order to do so, I will first present the evaluation and the program evaluation concepts, later on I will move on to exploring the two chosen classification forms upon which my own classification will be made. Afterwards, I will present my own narrative for the evolution of evaluation models, and finally I will present the evaluation models I considered to be the most important by using my classification form. The narrative that will be followed, will be based on combining the approaches of Guba & Lincoln (1989) and Alkin & Christie (2004, 2012) by structuring my own approach regarding the evolution of the program evaluation field.

3.1 The concept of evaluation

Just like the majority of developments in evaluation, the concept of evaluation itself is seen in many different ways, the main reason for this to happen is that the concept can be focused in a wide arrange of elements such as purposes, methods, and functions. However, all the uses of the concept have in common the notion of judging merit (Weiss, 1972). The most popular definition of evaluation is the one given by the American Evaluation Association which defines evaluation as a systematic process to determine merit, worth, value or significance (AEA, 2014). This definition has large similarities with the definitions given by well-known authors and organizations of the evaluation field such as Lincoln & Guba (1986; p.550), Scriven (1991: p.139), and the Joint Committee's (1994) definition of evaluation. Three important elements can be identified in this definition; value, merit and worth.

As stated by Stufflebeam & Shinkfield (2004), Guba & Lincoln (1986, 1989), and Mark, Greene & Shaw (2006), value, which is the root word of evaluation, denotes that evaluation essentially involves making value judgments. This in extent means that evaluation is not value free. This element is of great importance for the development of my thesis as it is directly linked to the establishment of judgments as well as recognizing the importance and influence of a set of values in the construction of these judgments. The way in which a situation is judged, determines the way in which it is framed, and in that sense, this will determine what and who will be taken into account in the evaluation design and

implementation. Additional to this consideration, the distinction between merit and worth is also relevant. According to Guba & Lincoln (1980), when referring to merit one needs to look at the quality of the evaluand, that what is going to be evaluated. This means that merit has a context free character dependant of the evaluand without taking into consideration what surrounds it. On the other hand, worth refers to the combination of excellence and service in an area of clear need within a specific context, meaning that when assessing the worth of an object, program or issue, this is being done taking into account the contextual elements surrounding it and which might be affecting it and may be affected by it as well. Taking into account this differences, an evaluand that has high merit might not be worthy.

I should now focus on program evaluation. I would like to explore the concept of program evaluation as being different from the one given to evaluation as a whole. The reason for doing this is that program evaluation only constitutes a type of evaluation, as we can also find evaluations directed towards personnel, processes, and outcomes without them being related to programs. Here I will present two definitions of what can be defined as program evaluation.

Patton defines program evaluation as

“The systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness and/or inform decisions about future programming (1997: p.23)”.

On the other hand, Rossi & Freeman (2004: p.29) define program evaluation as

“The use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social actions in ways that improve social conditions (2004: p.29)”.

It is possible to identify common elements in both definitions. For example both of them are directed towards assessing the effectiveness of a program by its improvement, and recognizing an evaluation to have a systematic character. However, the definition given by Rossi & Freeman (2004) can be seen as more comprehensive in the sense that it not only focusing on assessing the program but also considering the adaptive character the evaluation needs to have when taking into account the context and the political environment surrounding the program as these contextual elements do have an implication in the evaluation design. Given these definitions, I will ground my research on the definition given by Rossi & Freeman (2004) as it incorporates the context as an important element of program evaluation as well as aiming program evaluation towards improving social conditions. Both of these

elements are of interest for my methodological development, particularly because the context consideration is an important element for developing a systemic methodology, and because my methodological development is oriented towards social programs. This means that these types of programs are expected to have an impact and produce an improvement in the community in which they are taking place.

3.2 Evaluation Mapping

The story of how program evaluation has developed, can be told in several different ways, as many prominent authors in the field have made an effort for mapping the most relevant and most commonly used models in the field. An evaluation model as defined by Scriven (1991), is a term used to refer to a conception, approach or method of doing evaluation. It characterizes its author's view of the main concepts and structure of evaluation work (Stufflebeam , Madaus , & Kellaghan , 2000). Some examples of these mapping forms can be found in the work of Shadish, Cook & Leviton (1990), Rossi, Lipsey & Freeman (2004), Alkin & Christie (2004, 2012), Lincoln & Guba (1989), Worthen, Sanders & Fitzpatrick (1997), and Stufflebeam & Shinkfield (2007).

In order to narrate the historical development of these models, I chose to combine the approaches of Lincoln & Guba (1989), who describe the evaluation field development as having four generations, and Alkin & Christie's (2004, 2012) evaluation roots which classify evaluation approaches into three branches of an evaluation tree. Although it might be strange for the reader to use a combination of these approaches, this will help to get a clearer view not only of the approaches and the way they can be mapped but also of the principles and ideas on which each of these approaches were conceived. As it would be very extensive to describe in detail all the approaches accounted in the work of both authors, I will only explore the most important traits of most of the approaches, focusing extensively only on the Fourth Generation Evaluation which is the approach of central interest for this research and document. It is very important to note that although there are several ways in which evaluation evolution could be told, there is not a wrong or right way of doing so, as well as of choosing the most relevant of them. For this reason, choosing these two ways of viewing the development of evaluation approaches results convenient for the narrative I intend to have. Each of them will be described in what follows.

3.2.1 Mapping Methods

The work of Guba & Lincoln (1989) and Alkin & Christie (2004) will be briefly described in

what follows. At the end of the section I will present my own classification of evaluation approaches by taking into account the work of both sets of authors.

3.2.1.1 Evaluation Theory tree

As I have already stated, the evaluation theory tree corresponds to an alternative way among several options of organizing and mapping the most prominent evaluation approaches. Figure 7 and Figure 8 show the way in which the evaluation tree is organized and how the roots, branches, and leaves have a particular meaning in its distribution. Both trees have the same meaning and interpretation, however, the latter corresponds to an update of the first one as it was presented in a new edition of the book regarding the development of the Evaluation Theory Tree.

Figure 2. Evaluation Theory Tree (1st Edition)

Figure 3. Evaluation Theory Tree (2nd Edition).

The major differences between the two evaluation theory trees is that in the first one there were considered less approaches than in the second one, and most importantly, the tree foundations originally only considered as roots “social accountability” and “social inquiry”, in the latter version of the tree (See Figure 8), “epistemology” is added as a foundation of the tree, primarily as a foundation for the valuing branch.

When deciding what should be considered in the evaluation theory tree, Alkin & Christie (2004, 2012) found themselves in a position where they had to decide what was seen as theory. In that sense they focused on that which corresponded to evaluation models or approaches with a prescriptive character. Prescriptive modes are rules, suggestions, and guiding frameworks for good or proper evaluation (Alkin, 2003). They differ from descriptive models in that the latter are sets of statements and generalizations that predict or explain evaluation activities. However, they are not taken into account as there has not been a great deal of the kind of research on evaluation that might ultimately justify a descriptive theory. In that sense, the identification of an individual as a theorist and their writings as theories, implies that their work has been sufficiently developed and recognized as constituting a distinct approach (Carden & Alkin, 2012). Moreover, Alkin & Ellett (1985) maintain that all prescriptive theories must consider (a) issues related to the methodology being used, (b) the manner in which data are to be judged or valued, and (c) the user focus of the evaluation effort. These three main considerations are represented in the three branches

considered by the tree in the form of: use, methods, and judgment/valuing (see Figure 7 and Figure 8). Each theorist has been presented on the tree on a branch that represents his or her main emphasis among these tree branches.

According to Alkin & Christie (2004), the need and desire for accountability presents a need for evaluation. As a root for evaluation, accountability takes the broadest way possible. That is, accountability is not a limiting activity, but, rather, is designed to improve and better programs and society. It refers to the process of “giving an account” or being answerable or capable of being accounted for. Alkin (1972a) defines three types of accountability: goal accountability (appropriateness with which goals have been set), process accountability (adequate establishment of procedures for accomplishing the established goals), and outcome accountability (the extent to which the established goals have been achieved). The three accountability types are reflected in the evaluation writings.

The social inquiry root of the tree emanates from a concern for employing a systematic and justifiable set of methods for determining accountability. While accountability provides the rationale, it is primarily from social inquiry that evaluation models have been derived. Social inquiry, in its broadest sense, can be characterized as the systemic study of the behaviour of groups of individuals in various kinds of social settings by a variety of methods. It begins with the recognition that there is a unique social dimension to human action as opposed to merely a natural and psychological dimension (Alkin & Christie, 2004).

On the other hand, epistemology is the area of philosophy that deals with the nature and validity (or limitations) of knowledge. Those in evaluation who have historically and are currently engaged in epistemological discussions typically draw on one of three broad areas of thinking: positivism, constructivism (and related thinking), or pragmatism. Particularly, positivism is related to the evaluation approaches developed in the Methods branch, constructivism is related to those in the valuing branch, and pragmatism is related to those in the use branch. Thus, the basic axioms of these paradigms offer a broader framework for understanding the theoretical influences on evaluation theorists’ work. Epistemology serves as an important influence on evaluation theory more generally, albeit to varying degrees. Key evaluation concerns that are based in epistemological arguments include the legitimacy of value claims, the nature of universal claims, and the view that truth (or fact) is what we make it to be (Alkin & Christie, 2012).

The main branch of the tree is the continuation of the social inquiry trunk. This is the

evaluation as research, or evaluation guided by research methods, branch. This branch has been designated “methods” since in its purest form, it deals with obtaining generalizability, or “knowledge construction,” as Shadish, Cook, and Leviton (1991) refer to it. The authors also recognize that this branch not only refers to placing special attention to the methods used but also to emphasizing research methodology. This means, focusing not only on the methods used for conducting evaluation studies, but also on the techniques used with this purpose.

Another branch is the “valuing” branch. This branch emphasizes the vital role of valuing in evaluation. Theorists and models in this branch maintain that placing value on the subject of the evaluation (also known as the evaluand) is essential to the evaluative process. This branch was originally inspired in the work of Michael Scriven (1967) and Robert Stake (1967, 1975), for this reason their work is at the base of the branch. However, in contrast with the other branches, this branch is divided in two – objectivist and subjectivist – which distinguishes the two fundamental perspectives informing the valuing process. The objectivist influenced sub-branch to the left is most heavily influenced by Scriven’s views. He considered that it is the role of the evaluator to do that valuing. In essence, the important valuing role belongs to the evaluator. In the subjectivist sub-branch to the right, proponents argue that reality is an ongoing, dynamic process and a truth is always relative to some particular frame of reference. Thus, valuing must take place within the context of understanding the “subjective meaningfulness” of the evaluation information. In contrast to the objectivists, these theorists do not themselves solely make the valuing judgment (Carden & Alkin, 2012). Although all the branches derive from three foundation elements, this branch is particularly influenced by the epistemological foundation of the tree.

The third branch of the tree is “use”, which originally focused on an orientation toward evaluation and decision making. In essence, work by early theorists on this branch (Wholey, Stufflebeam, Guba) focused specifically on those empowered to use the information—most typically individuals who had contracted for the evaluation. Subsequent theorists have expanded the use concern to broader user audiences and to evaluation capacity building within the organization being evaluated. The three tree branches are not meant to be viewed as independent from one another, but rather have been drawn in a manner that reflects their relationship to each other (Alkin & Christie, 2004).

According to Alkin & Christie (2012), the three branches of the evaluation tree are related in the following ways. To the immediate right of the methods branch is the objectivist arm of

the valuing branch. Thus, the work of those theorists who are placed on the right side of the methods branch, leaning toward the valuing branch, reflect a secondary importance placed on valuing. Theorists on the left side of the valuing branch, closest to the methods branch, are primarily objectivist valuing theorists, with a secondary concern for methods. At this point an important consideration regarding the tree is that it has to be seen in a three dimensional way, this means that the far left side of the use branch will be closer to the far right of the valuing branch. Thus, the relational nature of the three branches helps to further depict theorists' approaches (Carden & Alkin, 2012). Taking this into account, moving to the far right side of the tree – the subjectivist arm of the valuing branch – are the theorists who reflect the relationship between valuing and use. These theorists are more concerned with individual stakeholder's actionable use of findings than are those placed on the left arm of the valuing branch. Correspondingly, the far left of the use branch reflects theorists whose primary concern is use but who have a secondary concern for valuing – in other words, a secondary concern for social justice and empowerment. The right side of the use branch reflects the relationship between use and methods. Thus, theorists on the right side of the use branch have a secondary emphasis on rigorous methods, whereas those on the left side of the methods branch place a greater emphasis on use than others who fall elsewhere on this branch. This description sums up the most important elements of the evaluation tree (Figure 7 and Figure 8). Alkin & Christie place the work of different authors in particular branches and positions.

3.2.1.2 The Four Generations of Evaluation

Guba & Lincoln's (1989), classification conceives evaluation as being developed through four generations of evaluation characterized as: measurement-oriented, description oriented, judgment-oriented, and negotiation oriented. The first generation evaluation or measurement-oriented evaluations corresponds to a time frame between the end of the nineteenth century and approximately 1930. This generation is characterized by being focused on developing ways to measure (tests) different aspects related to human differences in terms of learning, mental coefficients, effectiveness, and efficiency and how these elements could be improved, primarily focusing on applying tests in order to quantify the level of learning of school students. This generation, was majorly influenced by the legitimation given by the rise of social sciences at the time, and the emergence of scientific management movement in business and industry. It was characterized for ascribing a technical role to the evaluator; he or she was expected to know the full panoply of available instruments, so that any variable named for investigation could be measured. If appropriate instruments did not exist, the

evaluator was expected to have the expertise necessary to create them (Guba & Lincoln, 1989).

The second generation is known as the description oriented generation and it emerged as the result of the major deficiencies of the first generation. According to Guba & Lincoln (1989), after the World War I, it became evident that school curricula needed to undergo a dramatic revision, and an evaluation approach that could not provide other than student data (as the first generation did) could not serve the purposes of evaluation now contemplated. In this generation program evaluation emerged in the work of Ralph W. Tyler (his work will be described with more detail later). The second generation, is an approach characterized by description of patterns of strengths and weaknesses with respect to certain stated objectives. The role of the evaluator was that of describer, although the earlier technical aspect of that role were also retained. Measurement was no longer treated as the equivalent of evaluation but was redefined as one of several tools that might be used in its service (Guba & Lincoln, 1989). The second generation of evaluation can be placed in a time frame comprehending between approximately 1930 and 1942.

The third generation is oriented towards producing judgments. Just as the second generation emerged from the deficiencies of the first generation of evaluation (which was not explicitly called evaluation), the third generation evaluation emerged due to deficiencies in the second generation, primarily those related to the consideration of judgments. The call to include judgment in the act of evaluation marked the emergence of third generation evaluation, a generation in which evaluation was characterized by efforts to reach judgments, and in which the evaluator assumed the role of judge, while retaining the earlier technical and descriptive functions as well. This call, widely echoed in the profession, notably by Michael Scriven (1966), exposed several problems that had not been dealt with adequately in earlier generations. First, it required that the objectives themselves be taken as problematic, goals no less than performance were to be subject to evaluation. Further, judgment, as Stake (1967) pointed out, requires standards against which the judgment could be made. But the inclusion of value laden standards into a scientific and putative value-free enterprise such as evaluation, was repugnant to most evaluators. Finally, if there is to be judgment, there must be judge (Guba & Lincoln, 1989). After 1967 (year in which Scriven problematized the absence of the consideration of judgments in evaluation) a series of evaluation models started to be developed. All of these post-1967 models agreed on one point, however: judgment was an integral part of evaluation. All urged, more or less explicitly that the evaluator be judge

(Guba & Lincoln, 1989). However, there were differences in the extent to which evaluators were represented as appropriate judges.

Although each of the previously described generations represented an advancement from the previous one, Guba & Lincoln (1989) still considered that there were three major problems in the three generations that needed to be addressed, this is how they developed the Fourth Generation Evaluation also known as Constructivist evaluation. These problems were; a tendency towards managerialism, a failure to accommodate value-pluralism, and over commitment to the scientific paradigm of inquiry (Guba & Lincoln, 1989). A tendency towards managerialism refers to the typical relationship evaluator-client (the one who commands the evaluation), which follows a type of relation which is rarely challenged bringing as a result the disempowerment of other stakeholders as it is not possible to have inputs from others than the client, the manager (client) retains the right to determine if the evaluation findings are to be released, and this type of relationship is very likely to become one in which the evaluator and the client collude in choosing the appropriate way of conducting the evaluation.

Failure to accommodate value-pluralism refers to the fact that a society cannot be conceived as having a unique value set. Instead Guba & Lincoln (1989) view society as being value pluralistic. This is, as following several sets of values, however this also implies that a particular set of values will dominate in an evaluation, and understanding how value differences might be negotiated becomes an important issue. On the other hand, considering society as having a unique set of values is not the only problem. A second problem in evaluation is the fact that science is seen as value-free. The assertion of value-freedom for evaluations is completely resonant (reinforcing) with the managerial tendency already described. If values make no difference, then the findings of an evaluation represent states of affairs as they really are; they must be accepted as objective truths. The fact that the manager sets the boundaries and parameters for the study would then be relatively inconsequential, as would be the fact that he or she controls the questions asked, the methodology and the findings. The claim of value-freedom, is not tenable. And if that is the case, then the value-pluralism of our society is a crucial matter to be attended to in an evaluation. None of the evaluation approaches of the first three generations accommodates value differences in the slightest.

Third is the over commitment to the positivist paradigm. According to Guba & Lincoln

(1989), almost every approach developed in the three first generations of evaluation use the positivist paradigm to guide their methodological work, bringing several problems such as; considering the evaluand as existing in controlled conditions, this means that the context surrounding it and the relationship between them is not taken into account, an overdependence on formal quantitative measurements, the use of the scientific method closes out alternative ways to think about the evaluand, and because science is putatively value-free, adherence to the positivist paradigm relieved the evaluator of any moral responsibility for his or her actions.

All these elements are handled through the development of the Fourth Generation of Evaluation, as this is the focus of the document and as it does not corresponds to a generation per se (at least until now), this way of conducting evaluation will be explained with more detail when describing all the considered approaches.

3.2.1.3 Justification for these classification forms

Although I have already presented the classifications that will be used as a base to present the evaluation approaches in the current document, the need to justify this selection needs further explanation. The current document is not only focused, at least in theoretical terms, in evaluation but also on presenting the different theoretical developments and approaches of systems thinking that are relevant for the development of the research. For this reason, I sought to find some type of concordance in the narrative used for presenting both theoretical fields. In the particular case of systems thinking, the narrative used was based on the work of Midgley (1998, 2003) and Jackson (1990) who describe the evolution of systems thinking by using a three waves´ metaphor. I found this as being very helpful in terms that in my view it was a clear way of understanding the major changes that have taken place throughout the evolution of the field and also because they ease the comprehension of the reader and mine as a writer of the evolution of the topic.

Taking this into consideration, and the narrative of this document, a form of presenting the evolution of the evaluation field that was similar to the used to present the Systems Thinking theory was sought. Contrasting with the developments of the systems thinking field, the evaluation approaches are not as easily organized in a temporal frame, instead they are related more to the basic work of particular authors and the ideas they pursued. For this reason, I prefer to use guiding classification systems that not only had the theoretical bases for each of the approaches, but that could also set a guide in terms of a temporal scale to be

used. For this reason, the two approaches that fitted the most with my purposes were the Evaluation Theory Tree developed by Alkin & Christie (2004, 2012) and the Four Generations of Evaluation developed by Guba & Lincoln (1989).

The reader might be wondering why I decided to combine two classification systems instead of sticking up to one, there are some reasons for doing this. First, the four generations described by Guba & Lincoln are primarily focused on developing Fourth Generation Evaluation and not on giving an extensive detail on the particular developments of the previous generations. Besides, Fourth Generation Evaluation should not be seen as a Generation itself, but as a single approach. This to some extent, represents a challenge as it means that there is not too much available information regarding the particular approaches the authors considered when developing their classification. However, this set of generations aside of having a similar character to that of the “waved” approach used for the systems thinking field, also has a clear description of the dominant ideas and paradigms in each of the waves giving an idea of the evolution of the field through the generations. Second, the main evaluation model to be used in the current work is based on the Fourth Generation Evaluation approach. For this reason, taking into account the ground in which it was set was fundamental. Third, Fourth Generation evaluation was developed in the 1980’s. After that, several approaches with similar characteristics to the ones found in Fourth Generation evaluation have been developed. Therefore, they need to be taken into account as well.

On the other hand, the classification designed by Alkin & Christie is much more detailed than the one developed by Guba & Lincoln. The former is particularly designed with organizing and classification purposes that the Fourth Generations does not have. It results particularly important as the authors not only design the evaluation tree but when they had the opportunity they asked the authors they placed in their tree if they agree or not with their particular position and the assumptions attributed to their work. This is very important as in some way the classification has the direct approval of those placed in it. As well as the four generations of evaluation considered by Guba & Lincoln, Alkin & Christie also give relevance to the underlying paradigms of each of the branches in their tree.

When examining both classification forms, one can identify several common elements useful for developing a different form of mapping. The central common topic in both mapping forms starts with how valuing is taking place. This means that valuing takes place by two means: an epistemological component based on an objective or subjective way of valuing,

and who is responsible for doing such valuing (the evaluator, the client and the evaluator, or by involving stakeholders in different degrees for doing so). The relationship between these elements will shape the waves.

3.3 The proposed evaluation mapping for this research

I decided to use waves. Waves are very helpful not only for locating a progression given by time but also by noting a progression in the way in which issues of interest are being tackled in each one of them. The most important feature of addressing the evolution of the evaluation field by means of waves is that although a wave might come after another, there will always be remnants of the previous wave. This means that there is not a complete shift between waves but that there is a chained progression between them. Taking this into account, the features upon which waves will be described are: the way in which valuing processes are taken place within the approach, this is, the epistemological foundation of the approach, and who is in charge of doing such valuing. There is a correspondence between the nature of the valuing process and who is in charge of doing so in which a continuum between those in charge in making such decisions could be identified.

While Mertens & Wilson (2015) identify four paradigms in evaluation (post-positivism, constructivism, pragmatism, and transformative), Alkin & Christie (2004, 2012) recognize three paradigms (positivist, pragmatic, and constructivist). Taking this into account, as well as the four generations of evaluation described by Guba & Lincoln (1989) and a review of different evaluation models, I decided to work with the following four paradigms: positivist, post-positivist, pragmatic, and constructivist. Paying attention particularly to how values are managed, and how evaluators, clients, and other stakeholders are involved in the evaluative process.

At the heart of positivism is a belief that it is possible to obtain objective knowledge through observation and that such knowledge is verified by statements about the circumstances in which such knowledge is true (Stern , 2004). In the field of evaluation, House (1983) has discussed this tradition under the label of objectivism: ‘Evaluation information is considered to be “scientifically objective.” This objectivity is achieved by using “objective” instruments like tests. Presumably, results produced with these instruments are reproducible. The data are analysed by quantitative techniques which are also “objective” in the sense that they can be verified by logical inspection regardless of who uses the techniques.’ (House, 1983; p. 51). However, post-positivism on the other hand, represents an improvement from positivist

propositions in which it has incorporated fallibilism and falsificationism concepts (Adams , 2014). Post-positivists recognize that all observation is fallible and has error and that all theory is revisable assuming a critical position regarding our ability of knowing reality with certainty (Trochim, 2006). Hence, post-positivism believe that the goal of science is to hold steadfastly to the goal of getting it right about reality. In order to do so, it advocates just as positivists for the use of quantitative methodologies. However, it seeks to build a more complex research around it. It distinguishes itself from the different variants of positivism mainly through the view that the quantification and use of sophisticated statistical methods and mathematical models in itself and a priori do not enable the attainment of scientifically relevant insights (Adams , 2014). Values and biases are noted and accounted for, yet the belief is that they can be controlled within the context of scientific inquiry.

Pragmatism, embrace objectivity and subjectivity as two positions on a continuum and argue that deductive and inductive logic should be used in concert. Pragmatists do, however, move away from embracing equally the axioms of the post positivists and constructivist paradigms. Pragmatists are more similar to post positivists with regard to notions about external reality, with the understanding that there is no absolute “truth” concerning reality (Stanford Encyclopedia of Philosophy , 2013). More in line with constructivist thought, however, pragmatists argue that there are multiple explanations of reality and that at any given time there is one explanation that makes the most sense. In other words, at one point in time, a single explanation of reality may be considered “truer” than another. Furthermore, pragmatists, like post positivists, believe that causes may be linked to effects (Alkin, 2012). However, they temper this thinking with the caveat that absolute certainty of causation is impossible. Pragmatists are similar to constructivists in that they do not believe inquiry is value-free; rather, they consider their values important to the inquiry process.

Constructivists agree that there is no one reality, rather than several realities emerge from one’s subjective belief system. Hence, new knowledge and discovery can only be understood through a person’s unique and particular experiences, beliefs, and understandings of the world. These multiple realities are believed to be subjective and will vary based on the “knower”. In the inquiry process, the knower constructs a reality that is based on and grounded in context and experience (Alkin & Christie, 2012). Thus, there are no absolute truths, only relative truths. For constructivists, inquiry is considered value-bound, not value-free, and therefore, bias should be acknowledged rather than attempting to position the inquiry process so as to control it. Local relevance, then, outweighs and is of much greater

priority than generalizability. Cause and effect are thought to be impossible to distinguish because relationships are interdependent, and so simultaneously, all things have influence on all things.

The proposed four waves follow a paradigmatic order from positivism, moving on towards post-positivism, pragmatism, and constructivism. Each one of the proposed waves will be explained in what follows and examples of evaluation models will be describe in each one of them.

3.4 Evaluation Waves

This section is directed towards exploring the most prominent approaches among the evaluation models that have been developed and consolidated so far. These approaches will be presented in terms of waves.

The first evaluation wave does not properly corresponds to the practice of evaluation but setting up the foundations for doing so. As my mapping design is based solely on evaluation models, this wave will be represented by the work of Tyler who is seen as being the theoretical pioneer of the evaluation field as well as practical developments that took place in the United States.

3.4.1 First Evaluation Wave – Setting the ground for the formalization of the evaluation field

The first wave of evaluation follows a positivist paradigm. The particularity of this wave is that at this point, evaluation had not been established as a formal field but more as a practice carried out by different authors. However, their practice shared several traits and was influences by the ideas of efficiency, management and objectives which were very popular at the time particularly in educational evaluation.

The period comprised by the first wave includes the first half of the twentieth century. And is particularly influenced by the work of Fredrick Taylor and Ralph Tyler, who is considered as the father of the evaluation field. Taylor's work on scientific management became influential to administrators in education (Biddle & Ellena , 1964). Taylor's scientific management was based on observation, measurement, analysis, and efficiency which let to being smarter in the way work was done. These constituent elements made work and particularly evaluation to be objective-based. When applying such scientific methods to schools education has increasingly become characterized by standardized tests, expert opinion, and institutional

authority and control (Stoller 2015), primarily directed towards making inferences about the effectiveness of program curricula. The influence of scientific management principles is evident in the work of Tyler (1935, 1942, and 1950).

Tyler (1950) opined that instructional objectives should be specified in advance and used as a basis for determining and evaluating educational experience. Tyler's curriculum planning model became the popular framework that guided the planning and development of school curricular (Ireh , 2016). Although Tyler's work was directed towards conducting "evaluation" in education, several of his ideas were adopted to conduct and develop the field as a whole. Tyler originally coined the term evaluation to avoid the term "test" when assessing if students were learning. Some of his most important ideas are summed up in what follows:

"When thinking about evaluation, schools need to keep in mind the following key ideas: (a) education seeks to change behaviour of students in desirable ways; (b) These changes are the objectives of the school; (c) Evaluation should assess the degree to which these objectives are realized; (d) Human behaviour is too complex to be measured by a single instrument along a single dimension; (e) Evaluations must consider how patterns of behaviour are organized; (f) paper-and-pencil tests are not sufficient to appraise all the objectives; (g) The ways schools measure outcomes determine what is taught, how it is taught, what is learned, and how it is learned; (h) The responsibility for evaluation belongs to the school's staff and clientele; and (i) evaluation is an ongoing, cyclical process that should lead to the refinement of the objectives, the learning experiences that foster their attainment, and the outcomes attained (Madaus & Stufflebeam, 1989)".

These ideas are primarily related to the fact that evaluations need to take place in an objective based manner, it should use a wide variety of methods aside of what are traditionally seen as useful and it should be conducted with a shared sense of responsibility. For these reasons, these Tyler's ideas primarily gave rise to objective based approaches some of which will be explored in what follows.

On the other hand, the evaluation field also had a practical side from which it was developed, it is particularly related to the post war context of the United States, pointing at its rapid economic growth after World War II, the interventionist role of the United States federal government in social policy during the 1960s and the increasing number of social science graduates who became interested in policy analysis, survey research, field experiments and ethnography (Shadish , Cook , & Leviton , 1991). After the World War II, numerous major federal and privately funded programs were launched to provide urban development and housing, technological and cultural education, occupational training, and preventive health activities (Rossi, Lipsey , & Freeman, 2004a). As this happened, major concerns about the

setting up and execution of these programs started to take place. However, neither the public nor the private sector had the appropriate mechanisms, skills, and personnel to respond to the demand generated by these concerns, and for this reason graduate education in the field expanded to meet the needs of the day. Evaluation received the necessary institutional support for a new profession (public social programming), with a funding base (reimbursement for evaluative functions) and social legitimation (through government need for evaluation) (Austin , 1981). Without this support, evaluation might exist only as a small, applied academic discipline.

Although evaluation had the social legitimation of the government, at the moment of its establishment, it did not have a unique knowledge base. Instead, it heavily relied on existing methods and theories from all social sciences. Over time, evaluation theories changed and diversified to reflect accumulating practical experience (Shadish , Cook , & Leviton , 1991). Now the focus was not only on examining specific outcomes but also considering the possibility of examining other aspects of program implementation and with this came the development of different approaches of evaluation with varied purposes, methods, different levels of inclusion of stakeholders and the consideration of the context.

As could be seen in the description of the work of Taylor and Tyler, the first wave of evaluation had a positivist approach directed towards measuring how objectives and outcomes were being achieved as well as how institutions were performing. There have been several critics of Tyler scientific management model which led to further developments in evaluation. The Scientific Management Model predetermines the specific means in which learning will take place. In doing this, the model fails to recognize learning as a holistic process. Wilson and Cervero (1996) strongly critique the scientific management theory of program planning for believing that education planning and social and organizational context can be separated. On the other hand, Kliebard argues that the Tyler rationale is imperishable. In some form, it will always stand as the model of curriculum development for those who conceive of the curriculum as a complex machinery for transforming the crude raw material that children bring with them to school into a finished and useful product. (1970, p. 270). This statement is particularly related to standing against what is seen as a black box model. On the other hand, Kliebard (1970) also critiques the objective orientation of the model, as for him an adequate model for curriculum and instruction would start with the activity and not the statement of objectives. For Kliebard, the process of evaluation would be “one of describing and of applying criteria of excellence to the activity (Kliebard, 1970, p. 269)”.

These critics among others, led to the development of further approaches in evaluation.

3.4.2 Second Evaluation Wave

The second wave of evaluation, that is the post-positivist, is primarily concerned with developing adequate methods for conducting evaluations and particularly on designing and applying experiments. In this type of evaluations, the evaluator is in charge of all the decision making processes and the models are based primarily on quantitative methods. Stakeholders aside of the evaluator do not have any participation in the evaluation process primarily because evaluators are seen as having a superior knowledge to that of those involved in the program. For this reason, the methods they use, the interpretation they give to data, and the use given to the resulting information is seen as being the most adequate. According to Mertens (2015), methodological assumption in post-positivism reflect the use of scientific methods that allow the evaluator to discover laws about human behavior through empirical observations, prioritizing the use of experimental designs in the form of randomized control trials that require random selection of subjects and random assignment to intervention conditions. Finally, by being based on quantitative methods the approaches in this wave assume that valuation is value free as it is supported on facts. The work of Campbell & Stanley, Boruch, Cronbach, and Lewin & McEwan belong to this wave. They will be described in what follows.

3.4.2.1 Campbell & Stanley's Quasi Experimental and Experimental Designs

According to Campbell & Stanley (1963), there are many natural social settings in which the research person can introduce something like experimental design into his scheduling of data collection procedures (e.g, the *when* and *to whom* of measurement), even though it lacks the full control over the scheduling of experimental stimuli (the *when* and *to whom* of exposure and the ability to randomize exposures) which makes a true experiment possible. Collectively, such situations can be regarded as quasi-experimental designs. Quasi-experimental designs were developed to deal with the messy world of field research, where it is not always practical, ethical, or even possible to randomly assign persons to experimental and control groups (Alkin , 2004). As full experimental control is lacking, it becomes imperative that the researcher be thoroughly aware of which specific variables its particular design fails to control. Particularly, quasi-experiments consist on using controlled experiments where program evaluators randomly assign beneficiaries to experimental and control groups and then contrast the outcomes after the experimental group received a

particular intervention and the control group received no special treatment or some different treatments. What can result is valid knowledge about whether or not variation in that treatment package leads to variation in a package called outcome (Campbell D. T., 1969). This approach starts with questions and a methodology that address only a narrow set of program issues. Experimental and quasi-experiments are insufficient to address the full range of questions required to assess a program's merit and worth. The usual purpose of the controlled experiment is to determine causal relationships between specified independent and dependent variables (Stufflebeam & Shinkfield, 2007). Campbell and Stanley developed the basic taxonomy that distinguished the "true" or randomized experiments, the quasi-experiments, and the non-experimental design (Trochim , 1998).

Criticism regarding Campbell's work, is directed towards its epistemological foundation, its external, and its internal validity among others. According to Shadish, Cook & Campbell (2002), authors such as Kuhn (1962), and Harré (1981), direct their critique of the experimental and quasi-experimental designs towards the incommensurability of theories (theories are never perfectly specified for which they can be reinterpreted), the assumption that experimental observations can be used as truth tests , and the behavioral and cognitive inconsistencies between what scientists' do in practice and what scientific norms prescribe they should do. On the other hand, Cronbach (1982) and Kruglanski & Kroy (1976) critique the fact that experiments consider an atheoretical defined internal validity (A causes B) as being trivial without making any reference to constructs, as well as considering causations as impossible to determine in a single instance or experiment.

3.4.2.2 Boruch and the Randomised Field Experiment

The work of Robert Boruch is particularly similar to the work of Campbell & Stanley as he considers the randomized field experiment (or trial) as the ideal evaluation approach (Alkin, 2004). Boruch (1997) claims that any program (social, educational, or welfare) should be studied in a systematic manner by means of randomized controlled experimental methods, in order to gather valid and reliable evidence. However, not all programs are suitable to be evaluated by means of randomized experiments. Among the conditions that need to be considered to determine the appropriateness of a program to be tackled by these means are: the need for improvement in the present practice, the efficacy of the proposed intervention(s) must be uncertain, there should be no simpler alternatives, the results must be potentially important for policy, and the design must meet the ethical standards of both the researchers

and the service providers (Dennis & Boruch , 1989).

Using randomization, helps to assure that unknown influences on behaviour are equalized across the intervention groups. Although Boruch's work is very similar to quasi-experimental designs and observational studies, they have their differences as well. Quasi-experimental designs depend on assumptions or on methods other than randomization, to rule out competing explanations for treatment differences that may be uncovered, or to recognize bias in estimates of a difference in outcomes (Campbell & Stanley, 1966; Cochran, 1983; Cook & Campbell, 1979). However, part of the reason why quasi-experimental designs and observational studies have been useful is because it is not always desirable to employ randomized experiments.

According to Boruch (2003), when conducting randomized field trials, several elements need to be defined, among them are: the population, the units of randomization and the statistical power; interventions; the random assignment (this implies the impossibility of anticipating results); observation, measurement and theory; the variable associated with management; analysis of data; and reporting. Examples of the application of randomized field experiments could be found in the work of Burghardt & Gordon, 1990; Gueron & Pauly, 1991; and Mosteller et al., 1995.

Carey & Stiles (2016), find randomized field trials as being poorly suited for answering questions regarding why and how a program works as it does and most importantly they find major flaws in being able to find causal relationships between variables in randomized field studies. On the other hand, Deaton & Cartwright (2016), aside of critiquing the difficulties randomized field experiments have with causality, also critique the fact that as randomization moves oppressively over prior information from theory and from covariates, is wasteful and even unethical when it unnecessarily exposes people, or unnecessarily many people, to possible harm in a risky experiment. On the other hand, the non-random methods use prior information, which is why they do better than randomization. This is both an advantage and a disadvantage, depending on one's perspective. If prior information is not widely accepted, or is seen as non-credible by those we are seeking to persuade, we will generate more credible estimates if we do not use those priors. Finally, Jachimowicz (2015), highlights the need of randomized control experiments/trials for more sophisticated control groups, more careful consideration of other affected behaviors, and better conceptualizations of long term effects.

3.4.2.3 Lee Cronbach and the UTOS Development

Lee Cronbach's work did not solely derive from Campbell's work but instead it was inspired by Tyler's work as well as he was one of his students at the University of Illinois. For this reason. The main differences between his work and the previous experimental and quasi-experimental authors lies in the fact that Cronbach considers that there is no necessary conflict between experimental control and use of qualitative information or subjective interpretation, or between open-minded exploration and producing evidence that is objective. Cronbach's main interest on writing about evaluation emerged as a result of the dominant view in the evaluation field at the time, which he opposed to, as he stated

"...I argued that assessing the impact of a fixed plan is not the best use of evaluation; improving plans and operations is a higher goal, and to which a formal quantitative comparison usually contributes little (Cronbach , 1982)."

In this sense, he views evaluation as necessitating potential political accommodation. For this reason, he is more oriented toward evaluation utilization for enlightenment purposes than towards a concern for more immediate instrumental use. In Cronbach's view, evaluators should be educators rather than the philosopher-kings of Scriven, the guardians of truth of Campbell, or the servants of management of Wholey (Madaus, Scriven , & Stufflebeam , 1983). Cronbach et al. (1980, pp. 210) "advise the evaluator to be multi-partisan, to make the best case he can for each side in turn". This view goes in hand with what today is known as formative evaluation rather than the summative one, as he preferred to emphasize understanding principles, transfer of knowledge, and learning how to learn (Shadish , Cook , & Leviton , 1991). The central tenant about his view of evaluation, was directed towards being able to transfer processes between different sites, helping to design better future programs as

"The better and more widely the workings of social programs are understood, the more rapidly policy will evolve and the more the programs will contribute to a better quality of life (Cronbach et al., 1980, pp. 2-3)."

Taking this into account, Cronbach developed the UTOS model. Cronbach coins a set of symbols to define the domain of evaluation as a means of defining the manner in which evaluators might most productively enhance the use of enlightenment (Cronbach , 1982). This set consists of units (populations), treatments, observations (outcomes), and settings, the sample of this domain examined by the evaluator is referred to by the acronym "utos", the larger population from which the sampling took place is referred as "UTOS". For the design of the study, the central issue is to judge what are independent treatment units. The choice of

units of sampling follow from this, together with the strength of inference from u to U . The treatment specified for the unit is T while the treatment unit actually experienced is t (Madaus, Scriven, & Stufflebeam, 1983). Taking this into account, Cronbach et al., (1980, pp.314) state that what counts in evaluation is external validity, that is, the plausibility of conclusions about one or another UTO that is significant to the PSC.

3.4.2.4 Levin, McEwan and Cost-Effectiveness analysis

Cost-effectiveness analysis refers to evaluations that consider both the costs and consequences of alternatives. It is a decision oriented tool that is designed to ascertain the most efficient means of attaining particular educational goals (Levin & McEwan, 2003). The authors propose the necessity of considering both sources of information, effectiveness and costs, so that none of them separately will provide the wrong information to the evaluator.

Before moving on to describing the Cost-Effectiveness Analysis in more detail a differentiation between cost-effectiveness, cost-utility, and cost-benefit needs to be made. These three sets of concepts constitute three different analytical approaches to weighing the cost of an alternative against its outcomes. Although the three of them share one important aspect: their reliance upon cost estimates gathered via the ingredients method, which will be explained in more detail later on. Lewin & McEwan (2000, 2003) recognize several differences among them as well. Cost-effectiveness, compares two or more programs according to their effectiveness and costs in accomplishing a particular objective; it is very useful for evaluating alternatives that have a limited number of objectives. However, it does not tell us which alternative is worth the investment. Some examples of the use of cost-effectiveness can be found in the work of Dalgaard, Lewis & Boyer, 1984; Barnett, Escobar & Ravsten, 1988; Warfield, 1994; and Fuller, Hua & Snyder, 1994. In cost-utility analysis the evaluator gauges the overall utility of stakeholders, or their satisfaction with a number of alternatives. To measure utility, the evaluator attempts to elicit individual preferences for the relevant measures of effectiveness. These yields a set of utility “weights” that can be used to obtain a summary measure of utility for each alternative. Once this is accomplished, the analysis proceeds much like a cost-effectiveness evaluation. Some examples of cost-utility analysis use could be found in the work of Fletcher, Hawley & Piele, 1990; Lewis, 1989; Lewis, Johnson, Erickson & Bruininks, 1994; and Lewis & Kallsen, 1995. Finally, in cost-benefit analysis outcomes of several alternatives are directly expressed in monetary terms. This method can be used to assess directly whatever benefits outweigh costs, allowing a

statement about the program's desirability. Examples regarding the use of cost-benefit analysis can be found in the work of McMahon, 1998; Psacharopoulos, 1994; and Orr et al., 1996.

There are several elements that should be taken into account when estimating costs and measuring effectiveness, each of which will be analyzed in what follows. In the cost estimation there is a concept underlying called "opportunity cost". This cost makes reference to the cost or value one must give up when pursuing a particular intervention and not another. When estimating costs in a cost analysis with evaluative purposes, this opportunity cost is taken in terms of the value of the resources that were used or lost by applying them in one way rather than in another (Levin , 1983). To do this Levin & McEwan (2003) propose the Ingredients Method which relies upon the notion that every intervention uses ingredients that have value or cost. If specific ingredients can be identified and their cost can be ascertained, we can estimate the total costs of the intervention as well as the cost per unit of effectiveness, utility, or benefit. The steps to be followed in the ingredients method are the following: identifying the ingredients that are to be used (personnel, facilities, equipment, etc.), identifying adequate sources of information, and costing out the ingredients.

After costs are established, costs and outcomes must be jointly interpreted in order to rank alternatives from the most to the least desirable. Particularly, in cost-effectiveness analysis, the cost-effectiveness ratio of each alternative is obtained by dividing the cost of each alternative by its effectiveness. It's interpretation is given as the cost of obtaining an additional unit of effectiveness. Finally, one of the most common critiques of cost-effectiveness analysis is the fact that most of the ones which are found in literature are rhetorical claims with no data or incomplete analyses (Levin & McEwan , 2003).

As could be seen throughout the description of these set of models, they share common characteristics in the way they approach evaluation. For this reason, they are considered to follow a post-positivist approach. Particularly for their interest in developing experiments and models based primarily in quantitative methods, and based on the evaluators own judgment regarding the validity and the appropriateness of the results. However, as some critiques of these approaches show, the appropriateness of this methods is questions due to the theoretical validity in which they are founded, the lack of consideration of the context in which the evaluation is taking place, and the "objective" character given to the evaluation findings. For these and other reasons, the third wave of evaluation was developed.

3.4.3 Third Evaluation Wave

The third wave of evaluation moves forward from the second wave in that it no longer considers evaluation as necessarily having an objective character based solely on experiments and quantitative data. Instead, this wave constitutes a transition between objectivism and subjectivism, for this reason it is a pragmatic wave. Methods and the evaluators' opinion are not the only relevant issues to pay attention to, but clients start to be incorporated in different stages of the evaluation as well. Although the approaches considered in this wave do not explicitly consider a wide arrange of stakeholders, they do start to serve interests further than the one of the evaluator and sometimes those beyond the client's. In these approaches, valuing does not have an objective character and it can come from stakeholders different from the evaluator. However, it is usually represented by the interests of the client which are placed in a relevant position to put in practice the evaluation findings.

3.4.3.1 Rossi, Chen & Weiss and the development of Theory Driven and Theory Based Evaluation

Following the work of Cronbach, comes the work of Rossi, Chen and Weiss, the three of whom are representative figures of the Theory-driven evaluation, theory which according to Rossi is to a certain extent a fusion of Campbell's and Cronbach's approaches (Alkin , 2004). The work of Rossi and Chen is directed towards the development of what is known as Theory-driven evaluation, whereas the work of Weiss is directed to what is known as Theory-based evaluation. Although, both approaches have a very similar character, there are some elements that make them different, particularly those related to driving away from what is called "black box evaluation". The black box evaluation is the one concerned with the inputs and outputs of a program. The work of each of these authors will be explored in what follows.

Rossi views evaluation as an applied social research consisting essentially on the application of social research methods to provide credible information that can aid in the formation of public policy, in the design of programs, and in the assessment of the effectiveness and efficiency of social policies and social programs (Rossi, 2004b). He believes that the primary aim of evaluation is to aid stakeholders in their decision making on policies and programs. This means that evaluation findings are to be made available to stakeholders but they do not necessarily are critically influential in decision making. Evaluators should then try to be useful within the limits of professional ability and ethics. However, the primary goal should be to get it right and not getting along. Evaluation findings are information in the form of

judgments that are very helpful for decision making in each of five aspects of programs: needs assessments, program theory, implementation, impact and efficiency.

Chen, the other leading Theory-driven author, argues that the usefulness of program evaluation can be enhanced by achieving a balance between external and internal validation instead of maximizing only one of them. This makes reference to achieving scientific as well as stakeholders' credibility. For this reason, Chen (2004) argues that one of the major strategies that can achieve this features is a holistic assessment of a program under the framework of program theory. This is how Theory-driven evaluation emerged. The tenet of theory-driven evaluation is that the design and application of evaluation needs to be guided by a conceptual framework called program theory (Chen 1990, 2005). Program theory is defined as a set of explicit or implicit assumptions by stakeholders about what action is required to solve a social, educational or health problem and why the problem will respond to this action (Chen , 2012). Aside of seeking to assess if an intervention is working or not, this type of evaluation also seeks to answer questions regarding *why* and *how* it works as it does. Chen & Rossi state that black boxes programs are such because they have "no theory", goals are unclear, and measures are false, with the result that evaluations are "at best social accounting studies that enumerate clients, describe programs and sometimes count outcomes (1989, pp.299)". The black box is an empty box and Theory-driven evaluation should provide those programs with, among other things, good social science theory. But what is a "theory"? Various theory types have been suggested in the Theory-driven evaluation, the most popular of which is some form of causal model (Trochim & Cook , 1992).

Theory-driven evaluation should follow a set of steps: studying treatment; discussing stakeholders' and evaluators' views on outcomes; and examining why and how a program fares as it does (both normative and causal theories) (Stame , 2004). In order to do so, it uses two models: a change and an action model. According to Chen (2012), the change model describes the causal process generated by the program. The elements of a change model consist of goals and outcomes, determinants (focus which will clarify the design lines), and the intervention or treatment, this is any activity aiming at changing a determinant. On the other hand, an action model is a systematic plan for arranging staff, resources, settings, and support organizations to reach a target group and deliver intervention services. The action model consists of an implementing organization (the organization in charge of handling the implementation details), program implementers (those responsible for delivering services to the clients), peer organizations/community partners, intervention and service delivery

protocols (they state the nature, content, and activities of an intervention, ecological context (at a micro and macro level), target population. The action model should be implemented appropriately to activate the transformation process in the change model. For a program to be effective its action model must be sound and its change model plausible; its implementation is then also likely to be doing well. This conceptual framework of program theory should be useful to evaluators charged with designing an evaluation that produces accurate information about the dynamics leading to program success or program failure (Chen, 2012).

On the other hand is the work of Weiss (1997), directed towards developing a Theory-based (not a Theory-driven) evaluation. She addresses the “black box evaluation” problem in a different way from that of Rossi & Chen. According to Weiss, programs are confused because of the way decision making takes place; hence, insisting on aiming for “instrumental” use is pointless. Rather, evaluators should come to terms with the way decisions are made, and try to influence decision making in a more indirect way. Good program theory is not theory that is unaffected by politics, but a good theory of how politics works in practice (Stame, 2004). Hence, Weiss (1987) warns: evaluation is a rational activity in a political environment. Contrary to what Chen & Rossi believe, the black box is full of many theories; these are “theories of change”. Such theories of change take the form of assumptions and tacit understandings which could be used simultaneously on the same program (Weiss, 1995). Theories of change have two components; implementation theory, which forecasts in a descriptive way the steps to be taken in the implementation of the program, and the programmatic theory based on the mechanisms that make things happen.

3.4.3.2 Daniel Stufflebeam and the CIPP Model

The CIPP model is a comprehensive framework for guiding formative and summative evaluations of projects, programs, personnel, products, institutions, and systems. The model is configured for use in internal evaluations conducted by an organization's evaluators, self-evaluations conducted by project teams or individual service providers, and contracted or mandated external evaluations (Stufflebeam, 2004). Although the CIPP has shifted several times since it was initially designed, it emerged along with other conceptualizations, such as the ones developed by Scriven (1966), and Stake (1967) in the 1960's as one alternative to the views about evaluations that were most prevalent at that time, which were oriented towards objectives, testing, and experimental design (Stufflebeam, 1983).

The evaluation concept underlying the CIPP model and its checklist is that evaluation is

“A process of delineating, obtaining, providing, and applying descriptive and judgmental information about the merit and worth of some object’s goals, design, implementation, and outcomes to guide improvement decisions, provide accountability reports, inform institutionalization/dissemination decisions, and improve understanding of the involved phenomena (Stufflebeam, 2003)”.

In doing so, it explicitly uses four different types of evaluation directed towards serving four different functions and to answering different questions (see Figure 9): context, input, process, and product evaluation. Context evaluation involves identifying needs to decide upon program objectives, this means that it is oriented towards goals and seeks to answer questions related to getting to know what needs to be done. Input evaluation is oriented towards addressing plans, leading to decisions on strategies, funding proposals, and action plans chosen to be implemented. It seeks to answer questions of the “How something needs to be done?” type. Process evaluation is directed towards actions consisting of the identification of shortcomings in a current program to refine its implementation and to answering questions regarding if they are being take (or things are being done). And product evaluation measures outcomes for decisions regarding the continuation or refocus of the program (Stufflebeam, 2003; 2007) and assessing if it succeeded or not.

According to the CIPP model, evaluations should serve different clients in a way in which they present their audiences with evaluations that help promote a high quality administration, development, effective services, accountability, dissemination, and research. Summing all these elements up, CIPP is not directed towards proving but towards improving (Stufflebeam, 2003).

Figure 9. Key components of the CIPP evaluation model and associated relationships with programs.

Figure 9 (taken from the International Handbook of Educational Evaluation) depicts the relationship among the four forms of evaluation of the CIPP model with each of its core elements. The inner circle of the Figure, represents the core values that provide the foundation for one’s evaluations, these values may differ among the four forms of

evaluations. The circle surrounding the core values, is divided into four areas of interest corresponding to each type of evaluation. There is a bi-directional relationship between the areas of focus and each evaluation form. These bi-dimensional relationships are functional by grounding evaluations in the core values at the center of the scheme. Values provide the foundation for deriving the particular evaluative criteria. The criteria, along with questions of stakeholders, lead to clarification of information needs. These in turn provide the basis for selecting/constructing the evaluation instruments and procedures and interpreting standards (Stufflebeam, 2004). Evaluators and their clients must regularly employ values clarification as the foundation of their evaluation activities. Examples of the practice of evaluation based on the CIPP model can be found in the work of Zhang, Zeller, Griffith, Metcalf, Williams, Shea & Misulis (2011); Tokmak

3.4.3.3 Patton's Utilization Focused Evaluation

Utilization focused evaluation begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done from the beginning to end, will affect use (Patton, 2003; 2012). Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the focus in utilization-focused evaluation is on intended use by intended users. This focus allows evaluators to answer the question regarding whose values will frame the evaluation by working with primary intended users who have the responsibility of applying the evaluation findings and implementing recommendations. For this reason, a personal factor is key in utilization-focused evaluation. The personal factor is the presence of an identifiable individual or group of people who personally care about the evaluation and the findings it generates. Where the personal factor is absent, there is a marked absence of impact (Patton, 2003).

On the other hand, this evaluation model, does not depend on any particular evaluation content, model, method, theory or even use. It can include any evaluative purpose, any kind of design, and any kind of focus. It is a guiding framework as opposed to methodology which can include a wide variety of evaluation methods within an overall participatory paradigm (Ramirez & Brodhead, 2013). Utilization-Focused evaluation was originally designed to be carried out in 12 steps (Patton, 1978), however in the latest version of the approach Patton (2012) has added 5 new steps to it making it a 17 steps list to be carried out. These 17 steps can be summed up as follows (Patton, 2013): assessing program readiness, assessing

evaluator's readiness, identifying and engaging primary intended users, conduct a situational analysis, identification of primary intended uses, consider and build in process uses if appropriate, focus evaluation questions, check that fundamental areas for evaluation inquiry are being adequately addressed, determine what intervention model or theory of change is being evaluated, Negotiate appropriate methods to generate credible findings, Make sure intended users understand potential controversies about methods and their implications, simulate use of findings, data collection, organize and present data to be used, prepare an evaluation report, follow up with primary intended users, and meta-evaluation.

According to Patton (2003), there are several issues of concern which should be taken into account when conducting a utilization-focused evaluation: user responsibility and technical quality; user and most importantly primary users' turnover, which is seen as the "Achilles heel of utilization-focused evaluation"; building capacity for evaluation use; the variable evaluator roles linked to variable evaluation purposes, this means that different evaluation purposes require different roles for the evaluator; political underpinnings in this and other forms of evaluation; utilization-focused evaluators need special skills, they need to be politically savvy, skillful in group facilitation, able to decipher relevant internal organizational dynamics, and a user-friendly communicator (Torres, Preskill & Piontek, 1996); evaluation misuse, and ethics of being use-focused, this corresponds to the possibility of evaluators of becoming co-opted by stakeholders, given their close relationship and the use orientation of the evaluation.

3.4.3.4 Michael Scriven's Consumer Oriented Approach

According to Scriven (1967, 1991), evaluation can be defined as determining the value, merit and worth of things. This process identifies relevant values or standards that apply to what is being evaluated, performs empirical investigation using techniques from the social sciences, and then integrates conclusions with the standards into an overall evaluation or set of evaluations (Scriven, 1991). Scriven makes three key distinctions regarding evaluation: the goal of evaluation, the role of evaluation and the evaluator's responsibility. He focuses the goal of evaluation in judging value, the evaluator's main responsibility on making informed judgments and the role of evaluation as being a formative or a summative one. A formative role ascribed to evaluation is directed towards assisting in developing programs, whereas the summative role is directed towards assessing the value of the object once it has been developed.

Scriven, views evaluation as serving the interests of all those affected by the entity being evaluated not only the interests of a particular group. For this reason, his approach to evaluation is known as a consumer oriented approach. He argues that evaluands (the programs that are being evaluated) are good if they meet important needs, and that evaluators should look “for the need that is best matched by the actual performance and the performance that best matches some group of needs” (Scriven, 1981, p.130).

Scriven rejects the positivistic idea about reality being directly perceived through sensations and perceptions without mediating theories and without perceptual distortions (Feigl, 1969). Instead, he acknowledges that no single scientist ever observes “reality” completely and undistorted, but rather observes through biased perceptual filters (Shadish , Cook , & Leviton , 1991). For this reason, using multiple perspectives helps construct a more complete picture of reality; this is what he calls perspectivism. Scriven (1972, 1973, 1976, and 1983) acknowledges that there is a desire to minimize the many biases that can intrude when one is constructing value statements. However, he is aware of the impossibility of completely avoiding them, and instead encourages evaluators to be explicit about them and also to establish safeguards against unacknowledged bias.

The evaluator’s job, according to Scriven (1972, 1976), is to locate any and all program effects, intended or not, that might help social problems. Goals are a poor source of such effects. For this reason, one of the most important features of his approach to evaluation is advising evaluators to conduct “goal-free” evaluations. This type of evaluations, begin with the evaluators being blind to stated goals, and instead assuming the responsibility for determining which program outcomes to examine. Another essential element of his perspective, is the determination of a single value judgment of the program's worth (“good” or “bad”).

Scriven developed a Key evaluation Checklist as a means to plan and evaluate evaluations. Although it has a generic character, it can be adapted to be used in particular evaluations including evaluations of evaluations (meta-evaluations). This checklist contains relevant criteria for evaluating a particular object or making comparative evaluations of competing objects (Stufflebeam & Shinkfield, 2007). Scriven constructed such checklist to guide the collection of relevant evidence and to grade or rank the object of the evaluation. This checklist reflects his view that evaluation has multiple dimensions, must address pertinent values, should employ multiple perspectives, involve multiple levels of measurement, should

use multiple methods, and usually should culminate in a bottom-line evaluative conclusion (Stufflebeam & Shinkfield, 2007).

The last important element of Scriven's approach to evaluation is the consideration of a metaevaluation. According to Scriven (1968), evaluators have a professional obligation to ensure that their proposed or completed evaluation are subjected to competent evaluation. Meta-evaluation can take a formative or a summative orientation and can be conducted by using the Key Evaluation Checklist by assessing an evaluation as a product. In Scriven's view, a meta-evaluation is a way of controlling bias by making sure the evaluator get evaluated (Scriven, 1976; p.186). An example of this, is the fact that after data have been collected, analysed, and synthesized, the evaluator can bias the evaluation when writing the final report (Shadish , Cook , & Leviton , 1991). Scriven suggests making such biases explicit by making drafts of the report available to critics prior to final printing. Auditing a final report is a special case of metaevaluation – controlling bias by making sure “the evaluators get evaluated” (Scriven, 1976, p.176). Scriven recommends that evaluations be evaluated using his standard evaluation concepts and methods. To do so, evaluators can use his Key evaluation Checklist (1980, pp 113-116).

All these approaches are part of a pragmatic paradigm of evaluation as they are oriented towards getting to use the evaluation findings in practice. In order to do so, they involve the client in different stages of the evaluation. However, these approaches do not involve the perspectives of any other stakeholder. For this reason, the developments encompassed by a fourth wave of evaluative were developed. In these approaches, the stakeholders' interests, ideas and values are included and not only as a source of information but as a central element in developing the evaluation stages. They will be described in more detail in what follows.

3.4.4 Fourth Evaluation Wave

The fourth wave of evaluation moves a step forward in relation to the third wave as it acquires a subjective character in the sense that valuing is seen as coming from different stakeholders which need to be taken into account. Each of the stakeholders may bring to the fore different ways of viewing the same issue, as their views and forms of valuing are influenced by different elements. However, this does not means that quantitative and qualitative methods could not be used in the same intervention. Conceptions of reality are formed in a joint manner by those involved in doing so. This last wave includes Fourth Generation evaluation, however it is not the only approach in the wave as there are several

other approaches that place importance in the participation of stakeholders in the in the evaluation not only as a possible source of information but as active participants in planning, implementation, and analysis of results stages.

3.4.4.1 Robert Stake and Responsive Evaluation

Responsive evaluation (Stake , 2003) could be defined as a search and documentation of program quality directed towards addressing different stakeholder concerns and issues that may arise. In this particular type of evaluation the definition of issues by stakeholders is of particular relevance as it seeks to draw attention towards the interactivity, particularity, and subjective valuing already felt by those individuals related to the program. However, this responsiveness should not be mistaken with being responsive to program theory or stated goals of the program, as it is not the case.

This type of evaluation emerged as a way of responding to “preordinate evaluation”, which corresponds to a type of evaluation with a formal statement of goals, standardized tests of performance, particularly applied to students, value standards held by programs staff and the generation of reporting following a research journal form. This took place particularly because Stake et al. (1997) realized that disciplining impressions and personal experience led to better understanding of merit and worth than using needs to identify improvement with strict controls on bias. Having these differences with preordinate types of evaluation, it is very similar to other evaluation approaches such as the connoisseurship and the stakeholder evaluation, by sharing common characteristics such as drawing attention to social plurality of stakeholders, program activity and program uniqueness.

This evaluation model is directed towards discovering a program’s merit, worth, and shortcomings. For this reason, and by focusing on stakeholders’ issues and concerns, it recognizes multiple sources of valuing as well as multiple grounds. The role of the evaluator in this form of evaluation is directed towards inquiring, negotiating, and selecting particular issues and concerns from those recognized by the stakeholders, in order for them to be the organizational elements around which the study organizes. Aside of this, throughout the implementation of the evaluation, the evaluator must exchange descriptive data and interpretations with data givers, surrogate readers, and other evaluation specialists for recognizing misunderstandings and misrepresentations (Stake , 2003). However, it is the evaluator’s task to decide what methods will be used whether they are quantitative or qualitative. This selection depends partly on the situation being handled as the evaluation

must fit the “here and now”. Stake (2003) describes the role of the stakeholders as being more of sociological informants than that of subjects. This means that they are not seen as merely sources of data but they are also asked about what they see, how they feel, and to indicate the changes they have seen and not how they have changed. The feedback from responsive evaluation studies is expected to be attractive and comprehensive to the various groups, responsive to their needs. Responsive evaluation is not participatory evaluation, but it is organized partly around stakeholders concerns and it is not uncommon for responsive evaluation feedback to occur early and throughout the evaluation period.

According to Stake (2003), those who object to the responsive approach often do so on the ground that too much attention is given to subjective data. Placing value on the program is not seen as an act separate from experiencing it. The researchers’ own perceptions too are recognized as subjective, in choosing what to observe, in observing, and in reporting the observations. One tries in responsive evaluation to make those value commitments more recognizable. Objection to a responsive approach is also expressed in the belief that a single authority should specify the key questions. On the other hand, some examples of the use of a responsive approach in evaluation can be found in the work of MacDonald & Kushner (1982), Kushner (1992), McKee & Watts (2000), and Abma (2005) among others.

3.4.4.2 Elliot Eisner’s Educational Connoisseurship Evaluation

Educational Connoisseurship and educational criticism are two evaluation processes rooted in the arts (Eisner, 2003). Connoisseurship pertains to what Eisner (1991) has referred to as “the art of appreciation” for which its main goal is to engender an awareness of the qualities that constitute some process or object and to grasp their significance. Given its aim, it is characterized by having a qualitative character directed primarily towards raising awareness and creating consciousness by not only forming constructions about what one sees but also by taking into consideration the values that shape those constructions and the theories one uses in doing so.

The connoisseurship evaluation approach, is directed primarily to be used in educational realms. For this, Eisner (2003) argues that the importance of connoisseurship in an educational context of evaluation may be obvious in the sense that no one is ever in a position to make an evaluative judgment of an affair that is unaware of. Is in such awareness process in which individuals are able to appraise the quality of what has been experienced and consequently, connoisseurship is the art of appreciation. However, connoisseurship has a

private character, and for this reason for it to have real social consequences, it must transform into something with a public and sharable character, which in this case will be criticism (Eisner, 2003). The transformation of the content of connoisseurship into the public sphere through an act of criticism requires the “magical” feat of transforming qualities into a language that will edify others. It requires an act of representation. Effective criticism functions as the midwife to perception. It helps it come into being, then later refines it and helps it to become more acute (Eisner, 1998). In writing an educational criticism, Eisner (1976, 1985) identifies four generic dimensions: description, interpretation, evaluation (values are at its heart), and thematic. Some examples of the practice of educational criticism can be found in the work of Cohen, 1990; Kozol, 1991; Barone, 2000; and Kramer, 2015.

3.4.4.3 House & Howe and the Deliberative Democratic Evaluation

House & Howe (2003) propose the Deliberative Democratic form of evaluation as a tool for evaluators for solving the marked differences and conflicts that may arise among stakeholders. In that sense, this type of evaluation proposes a way of collecting, processing and analysing stakeholder perspectives in a systematic unbiased way, so that the perspectives could become part of the process of arriving at evaluative conclusions. This type of evaluation is based partly on experience with politicized evaluations in which stakeholders disagree and partly on philosophy and political science theory about deliberative democracy a theory that moves from simply aggregating stakeholder preferences to processes for transforming them (Gutmann & Thompson, 1996; Dryzek, 2000; Elster, 1998). In its name one can find two concepts as well, Deliberation and Democracy. According to House & Howe (2003), this evaluation approach aspires to be deliberative in two ways. First, evaluations inform people about relevant issues so they can discuss these issues in depth. Second, the evaluation study itself incorporates democratic procedures in its data collection and analysis. When evaluators encounter differing perspectives, they would do well to incorporate procedures that have served other democratic institutions. On the other hand, democracy requires that people discuss and deliberate on issues in order to come to a public understanding. During these discussions participants may change their view after being better informed. Participation in dialogue and deliberation is critical to arriving at conclusions.

The Deliberative Democratic evaluation approach emerged as a way of rejecting the sense of objectivity and impartiality promoted by positivists, as they strip away from the objective concept all value aspects leaving only facts. However, in House & Howe’s (2003) opinion

being objective means working towards unbiased conclusions through the procedures of the disciplines, observing the canons of proper argument and methodology. They deny that it is the evaluator's job to determine only factual claims and leave the value claims to someone else. As it happens with responsive evaluation, the evaluator is still in charge of the entire evaluation study however the stakeholder perspectives, values, and interests become an integral part of the study as well. One of the most important goals of including these perspectives is achieving an improvement in the validity of conclusions and increase the global legitimacy of the evaluation. Particularly because it seeks to be relatively unbiased in the study's value aspects.

Aside of considering these elements, there are three key principles in Deliberative Democratic evaluation (House & Howe , 2003): inclusion, dialogue, and deliberation. Although they may be described separately, in practice it is not that easy to separate them. First, inclusion refers to the mandatory inclusion of interests, values, and views of major stakeholders involved in the program or policy under review. Second, evaluators cannot presume to know how other people think without engaging in extensive dialogue with them. Having stakeholders discuss issues with each other can be part of the study itself or part of following up the study. Power imbalances are one of the biggest threats to democratic dialogues. These imbalances are endemic, and it is easy to see how they can disrupt and distort discussion. Evaluators should design evaluations so that relevant interests are represented and so that there is some balance of power among them. Third, Deliberation is fundamentally a cognitive process, grounded in reasons, evidence, and principles of valid argument. In the deliberative democratic approach, values are value claims (Beliefs) that are subject to rational analysis and justification. Perspectives, values, and interests are not taken as given or beyond question, but are subject to examination through rational process. Evaluation is a process for determining the validity of claims, claims that are emergent and transformed through deliberation into evaluation conclusions. Evaluation thus serves a deliberative democracy in which values and interests are rationally determined.

House & Howe main concern is directed towards addressing and changing the way in which facts and values are traditionally seen. Traditional evaluation, promotes the fact-value dichotomy which means that none of them can influence the other. However, according to House & Howe (1999; 2003) and House (1997), facts and values should not be seen as a dichotomy but a continuum in which they can influence each other in different degrees. In House & Howe (2003) view, objectivity and impartiality means that a statement can be true

or false, that it is objective or impartial if it is unbiased, that evidence can be collected from or against an statement's truth or falsity, and that procedures defined by the discipline provide rules for handling such evidence so that it is unbiased. Some examples of the practical use of Deliberative Democratic evaluation can be found in the work of Chelimsky (1998), Karlsson (1996), Greene (2000), Ryan & Johnson (2000), and MacNeil (2000). Other authors such as Stake (2000), Dryzek (2000), Nunnally (2001), and Stufflebeam (2001) express some concerns regarding the application of this type of evaluation with respect to several elements. First, is the fact that this approach might be bringing politics into evaluation, although House & Howe as well as other authors have made explicit the fact that an evaluative process is itself a political action. Second, aside of the Deliberative Democratic model is the "preferential" model of democracy which is most commonly used. However, regarding this issue, House & Howe consider their approach as being better than the other given the possibility of changing initial stakeholders' preferences. Third, is deliberative democratic evaluation too idealistic to be implemented? Its attempt to make evaluation just, for attaining democratic participation, incorporating the views of the interested parties, and having an ethical component represents a real attempt for improvement.

3.4.4.4 Guba & Lincoln and Fourth Generation Evaluation (also known as Constructivist Evaluation Approach)

According to Guba & Lincoln (1989), Fourth Generation Evaluation emerged as a need to move from the governing positivist paradigm from the first three evaluation generations. Positivists assume the existence of an objective reality, with an epistemological character corresponding to the observer-observed duality, which means that the observer has the chance of achieving an objective view of that which it is observing by appreciating it from the outside without influencing it or being influenced by it (Guba & Lincoln, 1994). Moving from this paradigm, implied moving from focusing on the facts, namely a scientific approach, to an approach that includes the human, political, social, cultural, and contextual elements that are involved in the evaluation, and most importantly on dismissing the observer-observed duality and replacing it by shared constructions that are influenced by social and cultural factors. In order to do so, Guba & Lincoln (1989) based their work in two different approaches; the responsive form of evaluation developed by Stake (2003), and constructivism which is one of the key concepts of the interpretive paradigm. The responsive character of the new form of evaluation was directed towards focusing the evaluation design, primarily in determining which parameters and boundaries are to be considered in the evaluation, whereas

constructivism constitutes the actual methodology of conducting the evaluation.

The Responsive character of the Fourth Generation Evaluation

According to Guba & Lincoln (1989), an evaluation must begin by determining what questions are to be addressed and what information is to be gathered. These elements are known as “advanced organizers” and they are used to clarify the variables, objectives, and important decisions that need to be addressed by the evaluation.

As explained in the section referring to Responsive evaluation, this type of evaluation has its own advance organizer as well. These are the claims, concerns, and issues about the evaluand (the subject of the evaluation) that are identified by stakeholders, that is, persons or groups that are put at some risk by the evaluation. Guba & Lincoln (1981), identify three broad classes of stakeholders; agents, beneficiaries and victims. Agents are the individuals involved in producing, using, and implementing the evaluand. Beneficiaries are those who profit in some way the use of the evaluand. And the victims are the ones who are negatively affected by the use of the subject of evaluation. The responsive character of this type of evaluation is not only given by the fact that it considers different stakeholders views, but also because it responds to the collection of information that takes place by doing so. However, it is not only focused on collecting the information, but also on conducting the evaluation in such a way that each group must confront and deal with the constructions of all the other groups (other stakeholders), by following a process called hermeneutic dialect (this will be explained in more detail later). As each group copes with the constructions posed by others, their own constructions alter by virtue of becoming better informed and more sophisticated (Guba & Lincoln, 1989).

Responsive evaluation has the ideal of reaching consensus on the claims, concerns, and issues among the stakeholders. However this is rarely possible, and that is where conflicts arise. The evaluator plays an important role in solving these conflicts as it plays a central role in the collection of new information necessary to solve them, and also because it play an active role in preparing and moderating an agenda for negotiation with representatives of the relevant stakeholders. The result of such negotiation process(es) are arrived jointly by those involved in them. The stage is set for recycling the evaluation. Such iteration and reiteration is typical of responsive evaluation; evaluations are never complete but are suspended for logistical reasons, such as the timing of a mandated decision or because resources are exhausted (Guba & Lincoln, 1989).

The constructivist approach of the Fourth Generation Evaluation

Guba & Lincoln (1989) propose constructivism as the dominant paradigm of the Fourth Generation Evaluation. A paradigm as defined by Patton (1978, pp. 203) “*is a world view, a general perspective, a way of breaking down the complexity of the real world.*” It is not possible to prove or disprove a paradigm in an absolute sense. However, Guba & Lincoln (1989) consider that the constructivist paradigm is the appropriate one to extend their work, given its ontological, epistemological, and methodological characterization.

Ontologically, constructivism denies the existence of an objective reality (the objective reality was at the core of positivism, the previous dominant paradigm of evaluation). Instead as Jonassen (1991) describes it, constructivism asserts that reality is in the mind of the knower, that the knower constructs a reality, or at least interprets it, based upon his or her apperceptions¹. However, it does not preclude the existence of an external reality; it merely claims that each of us constructs our own reality through interpreting perceptual experiences of the external world (although clearly many constructions will be shared). This means that constructivism follows a relativist reality.

The relativist posture of the constructivist ontology places a constraint in its epistemology as it is impossible to separate the inquirer from the inquiry, this means that there is an interaction between the observer and the observed that creates what emerges from an inquiry. Furthermore, as Guba & Lincoln (1989) state, an inevitable element in the inquiry mix is the values of the inquirer and a whole variety of other persons in and around the inquiry; The values of the inquirer, which in the particular case of the Fourth Generation Evaluation is the evaluator, inevitably enter the inquiry in connection with the whole series of decisions involved in designing, mounting, and monitoring. The values of the involved individuals (which in this case are the stakeholders), also exert influence as the inquiry takes place in value contexts. Values have a fundamental role in constructivism as they are reflected in the theory that may reinforce the inquiry, in the constructions that the inquirer and others bring to the inquiry, and in the inquiry paradigm itself. Moreover, all of these values may be disjunctive – for example, there may be a value conflict between the theory undergirding an inquiry and the paradigm that guides it. For all these reasons, say the constructivists, values cannot be ignored; their very influential role in all inquiry must be acknowledged.

¹ According to the Oxford Dictionary the term “Apperceptions” refers to the mental process by which a person makes sense of an idea by assimilating it to the body of ideas he or she already possesses.

On the other hand, and as a consequence of the ontological (the relativist reality) and the epistemological (interaction between stakeholders and evaluator) assumptions of constructivism, methodologically it rejects the controlling, experimental scientific approach and substitutes it for a hermeneutic/dialectic process that takes explicit advantage of the interaction between the observer and what is observed in order to create a constructed reality at a particular time. This dialectic process, allows the critique of the constructions of the different parties, as well as providing opportunities for revised or new constructions to emerge. This dialectic character leads to a better understanding as it works by the iterative interaction of those involved in the dialectic process, namely the evaluator and the stakeholders. This process follows a hermeneutic character in that it aims at developing improved joint constructions, and it is dialectic in that it involves the proximity of conflicting ideas, forcing those involved in the process to reconsider and reflect upon their previous positions.

According to Guba & Lincoln (1989) the methodology followed by the constructivist paradigm has a series of caveats as well as of entry conditions. Its caveats are related first to the simplistic consideration of any complex methodology, this means that it results extremely difficult to capture the essence of non-linear processes by means of a two-dimensional flow chart. Second, and most important of all, the constructivism methodology was not simply developed as a reaction to the positivist form of inquiry. Instead it represents a strong alternative position generated from its own assumption.

Following these caveats, Guba & Lincoln (1989) state that there are four entry conditions for the constructivist inquiry. First, the study needs to be pursued in a natural setting particularly at the same time/context frame that the inquirer seeks to understand, this makes particular refers to the importance of the context in the generation of constructions. Second, constructivists assume that they do not have enough a priori knowledge about the context frame surrounding them, because of this assumption, they do not predefine what questions to ask. What is needed under these circumstances is a highly adaptable instrument that can enter a context without prior programming. The human is the instrument of choice for the constructivist, and, it should be stressed, the only possible choice during the early stages of an inquiry. Objections that humans are subjective, biased, or unreliable are irrelevant, for there is no other option. Third, as humans are the central instrument of the constructivist inquiry, the preferred methods are the qualitative ones. However, there is nothing against the use of quantitative methods as they can be used when it is appropriate to do so. Fourth,

constructivists insist on the right to use tacit knowledge (Polanyi , 1964). Polanyi (1964, 1966) defines tacit knowledge as what we know but that we cannot tell; it includes so much embedded learning that its rules must be impossible to separate, thus it is impossible to reproduce in a document. That is, tacit knowledge normally cannot be spoken, but rather demonstrated and imitated. When these four specifications are met, it is possible to mount a constructivist inquiry. The process used in the next stage is the hermeneutic dialect.

The hermeneutic dialect process of the Constructivist Inquiry

The hermeneutic dialectic process involves four interacting elements that follow a cyclical and recycling process until consensus (or not) is achieved. This means that the process will stop when consensus is being achieved or until the given conditions make it mandatory for it to stop. The first element is the selection of those who will enter the hermeneutic process. In this case, the way of selecting them is by “Purposive sampling”. Patton (1980), described six different types of purposive samples aside of having a representativeness or randomness purpose, these are; sampling extreme or deviant cases, sampling typical cases, maximum variation sampling, sampling critical cases, sampling politically important or sensitive cases, and convenience sampling. The best of these purposive samples that adapts to constructivism is the “maximum variation sampling”. According to Patton (2002), consists of purposefully picking a wide range of cases to get variation on dimensions of interests. This type of sampling, documents diverse or unique variations that have emerged in adapting to different conditions, as well as identifying important common patterns that cut across variations. It is the type of sampling that provides the broadest scope of information.

The second element, has to do with the continuous interplay of data collection and analysis that occurs as the inquiry proceeds. The third element, has to do with grounding the findings that emerge in the construction of the respondents. As data collection, analysis and critique take place at the same time, a joint constructions begins to emerge about which consensus (if possible) can begin to form. This joint construction differs from the individual constructions originally offered by respondents and evaluators. However, it is grounded in all those constructions, derived from them via the hermeneutic dialectic process. These joint constructions must work, be relevant, and exhibit a modifiable character (Glaser & Strauss, 1967; Glaser, 1978). The fourth element corresponds to the emergent design, this means that as the evaluation keeps gathering information and addressing constructions, it can focus more the design of the inquiry (in this particular case it would be an evaluation) as it can identify

its most important elements.

Summing up, the hermeneutic dialectic process has at its core the achievement of joint constructions, and the ideal of reaching consensus in one or multiple constructions. However, if consensus is truly achieved, there is still space for further inquiry; as the appearance of new information may imply further reconstructions and interpretations. On the other hand, if consensus is not achieved, its lack will point out the points between which differences appear, points for which further negotiation, and reconstructions might be needed. Such joint constructions, must have an emic and an etic component; this corresponds to considering the insider and outsider views which can only be reached by expressing own's points of views, and exposing them to the interpretation and critique of others. Judgments of that joint construction's fit, work, relevance, and modifiability must be made by inquirer and respondents jointly. Thus design, emergent theory, and "findings" will all represent a unique combination of inquirer and respondent values and judgments.

As can be seen, the focus of the constructivist paradigm, that of consensus, differs from the traditional scope on the criterion of correspondence between findings and reality of the positivist paradigm. This difference resides primarily in the different appreciations both paradigms have of the observer/observed duality. As constructivism considers the existence of multiple realities, the correspondence sought in the positivistic paradigm cannot be achieved as it can only be achieved by considering a single objective reality.

The final result of the iterations of the hermeneutic/dialectic circle is a case report, it represents the joint construction that emerges as a results of conducting this process. One of the main purposes of this case report, is that the reader may come to the realization not only of the state of affairs that are believed by those who make the constructions, but also the underlying motives, feelings, and rationales underlying them. As can be seen, the methodology of the constructivist is very different from that of the conventional inquirer. The latter is linear and closed. By contrast the former is iterative, interactive, hermeneutic, at times intuitive, and most certainly open (Guba & Lincoln, 1989).

The Methodology of Fourth Generation Evaluation

As I previously stated, Fourth Generation Evaluation is a combination of responsive focusing, by using the claims, concerns, and issues of stakeholders as the organizing elements, and the constructive methodology, aiming to develop judgmental consensus among stakeholders who earlier held different, maybe conflicting, emic (inside, this means individual) constructions

(Guba & Lincoln, 1989). The specific methodology of the Fourth Generation Evaluation, includes 12 steps, which although they appear to be linear in sequence, their design and application is not rigid. Rather, they indicate progression only in a general way, as back and forth movements between the different steps is not only possible but desirable as well. Each of the general steps will be indicated and described in what follows.

1. Initiating a contract with the client or sponsor commissioning the evaluation: this contract implies the identification of the client, the identification of the evaluand (what is going to be evaluated) which is based on the different interpretations and constructions of those involved, particularly having as a benchmark the interpretation of the client, and the establishment of purpose for evaluation. According to Guba & Lincoln (1981), there are four possible ways of defining the purpose of evaluation: it could be a formative merit evaluation, a formative worth evaluation, a summative merit evaluation, or a summative worth evaluation. Guba & Lincoln (1989) describe these types of evaluation as follows. A formative merit evaluation is one concerned with assessing the intrinsic value of some evaluand with the intent of improving it. A formative worth evaluation is one concerned with assessing the extrinsic value of some evaluand with the intent of improving it. A summative merit evaluation is one concerned with assessing the intrinsic value of some evaluand with the intent of determining whether it meets some minimal standard for modernity, integrity, and so on. A summative worth evaluation is one concerned with assessing the extrinsic value of some evaluand for use in some actual context of application. This step of the evaluation also includes specifying other type of information related to the commitment of the client (the one who commands the evaluation), specifying the type of methodology that will be followed in the evaluation, specifying the type of access to information, documents, data and previous reports, confidentiality, and other elements that will be important to establish and initiate the evaluation.
2. Organizing to conduct the evaluation: this element correspond to a series of elements that could be problematic in further stages of the evaluation if the evaluator does not take them into account since the beginning. Particularly, Guba & Lincoln (1989) refer to four elements: selecting and training the evaluation team, which can provide a wide arrange of perspectives and expertise; gaining entrée and getting established, which is the base for consolidating trust; making logical arrangements. These arrangements have to do with the logistic of the different activities that are planned to be carried out

throughout the evaluation; and being aware of the local social, political, and cultural factors surrounding the evaluand and the evaluative process, these factors are fundamental for conducting the evaluation and getting a more complete perspective about the dynamics of the program which might lead to a more successful evaluative process.

3. Identifying stakeholders: The stakeholder concept is widely seen by different authors (Thompson, 1967; Freeman, 1984; Guba & Lincoln, 1989) as the group of individuals who have a stake in an organization, project or inquiry. This is, any group of individuals who can affect or who are affected by the achievement of the objectives of the organization, program, process, or inquiry. For the particular case of Fourth Generation Evaluation, Guba & Lincoln (1981) specifically identify three types of stakeholders: agents, beneficiaries and victims. One of the central tasks of the evaluation is to identify the adequate stakeholders for conducting the evaluation. However, this task does not corresponds solely to him or her. Instead, the inclusion or exclusion of stakeholders must be determined by negotiation, in which each audience that wishes to do so can present its case from a position of equal power. The role of the evaluator is to arrange the situation where this negotiation can take place. If audiences are not sufficiently knowledgeable or sophisticated at the start, it is the evaluator's moral obligation to take steps necessary to bring them to whatever minimal level is required for full participation. When negotiations have been entered into with identified stakeholders, agreements must be formalized.
4. Developing within-stakeholder-group joint (collaborative, shared) constructions via the hermeneutic/dialectic circle process, focusing specifically on claims, concerns, and issues. A claim is any assertion that a stakeholder may introduce that is favorable to the evaluand. A concern is any assertion that a stakeholder may introduce that is unfavorable to the evaluand. An issue is any state of affairs about which reasonable persons may disagree. Different stakeholders will harbor different claims, concerns, and issues; it is the task of the evaluator to bring them to light and to address them in the evaluation (Guba & Lincoln, 1989). This step involves the first application of the hermeneutic dialectic circle, its main purpose is to uncover as many different constructions as possible by being focused on a particular group of stakeholders and seeking to address how they view the evaluand by means of their concerns, claims, and issues. There is no guarantee that a joint construction encompassing everyone's input can be attained, but consensus is the goal. When there is no consensus,

competing constructions can be carried forward; it is on matters of non-agreement that Step 7 and those following are based. As any hermeneutic dialectic process, one's constructions should be put under critique of the others so that joint and enriched constructions can be achieved. The process may come to an end after achieving redundancy, consensus, or irreconcilable differences.

5. Testing and enlarging within-group constructions: this is achieved by introducing new or additional information and by enabling group members to achieve higher levels of sophistication in dealing with such information. Again, the main goal is being able to achieve consensus, however as new concerns, claims and issues will be introduced, this is less likely to happen than in the previous step. Contrary to this, the points of difference are those on which the evaluation will be focusing, and those in which consensus was achieved will be dropped. Documents and records, observation, professional literature, other stakeholder circles, and the evaluator's own constructions are sources of new information. When all this information has been exposed and dealt with, the end of the hermeneutic interaction is essentially reached.
6. Sorting out resolved claims, concerns and issues: these claims, concerns, and issues are those in which consensus has been achieved and can be of historical interest as well as being considered in other settings.
7. Prioritizing unresolved claims, concerns, and issues: in some cases the evaluation might not be able to address all the unresolved claims, concern, and issues as they might be too many or because the evaluation resources are too narrow to work with all of them, for this reason prioritization of issues needs to be achieved. This prioritization task should be carried out in a participatory way that provides opportunity for input and control to each of the stakeholding audiences. A possible way of doing so, is by conducting a new hermeneutic dialectic circle with representatives of each of the stakeholder groups, however it does not necessarily need to be conducted in this way.
8. Collecting information referring to the unresolved claims, concerns, and issues. The unresolved claims, concerns, and issues correspond to those constructions that could not achieve a joint or consensual character in the previous hermeneutic dialectic circles. For this reason, the evaluator must collect information that will, when added to existing constructions, lead to reconstruction, and the determination of ways and means to make appreciation of that new information possible. This new information

can come from different sources such as documents, previous experiences of the evaluator, tests, and others.

9. Preparing the agenda for negotiation: According to Guba & Lincoln (1989), it is useful to think of this step as analogous to the more usual drawing up of a set of conclusions and recommendations. Conventionally, this is the task of the evaluator because he or she is the expert, is detached from and disinterested in the local setting (“objective”) and possesses the requisite technical expertise to deal with data and their interpretation. But such a position is intolerable within constructivist assumptions. Data are never value-free, as we have seen, and they certainly do not “speak for themselves”, that is, they do not carry their own meaning and interpretation in an unequivocal way. Evaluators, guided consciously or unconsciously by their own values, select which facts are to be determined and once obtained, what they mean. If that is the case, as constructivists, then the evaluator’s values are very much at issue; they ought not to be solely determinative, especially in a pluralistic, multivalued society. The negotiation session, representing the full array of stakeholders, is a mechanism intended to display and take account of the full range of values. If the earlier hermeneutic circles are determinative of which “facts” should be collected, this negotiation session decides what they will mean.
10. Carrying out a negotiation: in the Fourth Generation Evaluation, negotiation is expected to be carried out by using a Hermeneutic Dialectic circle. However, it is not necessarily mandatory to use it in every case. The negotiating circle (following a Hermeneutic Dialectic process) is selected in ways that follows the principle of participatory (empowering, educative) evaluation, but without being merely representative in the statistical sense. Instead, those representatives must be conscious of the constructions of the stakeholders they represent (without being instructed). Given that the claims, concerns, and issues have been addressed, three possible results may appear: full resolution, which implies a direct evident actions about that which has been addressed; partial resolution, this means that some of the claims, concerns or issues being addressed could not be resolved but others did, and no resolution which means that none of that which was addressed could come to resolution.
11. Reporting: in this type of evaluation reporting always comes by means of generating a report, this report delineates the joint construction(s) that has emerged via the hermeneutic dialect process(es) that have been conducted in the previous stages of the

evaluation. It should be as explicit as possible so that the reader could understand how, and why constructions were made.

12. Recycling: in this type of evaluation there are almost always unresolved claims, concerns, and issues which might be brought about by the appearance of new information or increased sophistication. For this reason, they might be addressed again so that they might be solved and most likely raise new claims, concerns, and issues that were not considered previously.

These are the 12 steps of the Fourth Generation Evaluation process, if the reader wants to know about them in more detail, it should remit to the original work of Guba & Lincoln (1989) on the development of the Fourth Generation Evaluation.

The role of the evaluator in the Fourth Generation Evaluation

Throughout the different approaches of program evaluation the role of the evaluator takes different roles. For example, although Fourth Generation evaluation sets its ground in responsive evaluation, there are substantial differences in the role given to the evaluator in both of the approaches. In general, in most of the program evaluation approaches, the evaluator is seen as an individual or group of individuals (for the case where there is an evaluation team), with superior knowledge only comparable to the one of the client of the evaluation as both of them are the ones who make the important decisions about how to approach the stages of the evaluation.

However, in the case of Fourth Generation evaluation, the evaluator does not have a messianic figure with superior knowledge who comes to solve the program's problems as it might have in other evaluation approaches. Instead, aside of conducting several stages of the evaluation on its own, the evaluator also has a role of facilitator for conducting some stages of the evaluation, as well as for easing the communication and joint work between the stakeholders. In the case of the hermeneutic/dialectic circle, the evaluator is in charge of setting up the appropriate conditions to carry out the dialectic process, these conditions might not only be related to the accommodation of the physical space and the availability of appropriate tools, but also to help those participating in it in exploring relevant issues, concerns and claims.

Critique of Fourth Generation Evaluation

Gregory (2000) conducts a critical analysis about the participative approaches in evaluation theory, one of which is Fourth Generation Evaluation. According to Gregory (2000), there are

three types of barriers in its ability to achieve participation. The first barrier corresponds to a structural barrier. Given that evaluation has a political connotation, Guba & Lincoln seek to reduce the influence of this political character by means of establishing conditions for conducting the dialectic process. However, in doing so, they allow the client to ‘claim a legal need to retain veto power over decisions’ and, in so doing, privilege the rights of the client, which effectively supports a structural barrier (Laughlin & Broadbent, 1996). Guba and Lincoln seek to overcome this by stipulating that the evaluator is morally obliged to bring parties which do not possess sufficient knowledge or sophistication to the minimal level required for full participation. However, this brings new problems as a “minimal level” is not defined. Guba and Lincoln go on to propose that the evaluator may act as an advocate or introduce third parties to act as advocates in order to redress imbalances in knowledge between stakeholding parties. This introduces a set of new problems to the evaluation process since their consideration of the role of advocates is underspecified. Whilst Fourth Generation Evaluation may appear to take seriously the impact of power relations and does seek to alert users of the methodology to the need to consider and involve stakeholders, it does not go on to discuss in necessary detail the range of values, motivations, prejudices, etc. which will emerge and need to be effectively managed in the participatory environment (Gregory , 2000).

A second barrier, corresponds to an administrative barrier related to the gatekeeping role of the evaluator. Guba & Lincoln state that formal gatekeepers have authority and informal gatekeepers have influence, but either has the power to support or hinder an evaluation’ (1989:198). , they advise the evaluator of the endemic nature of gatekeeping practices and recommend that each gatekeeper may require a particular *quid pro quo* to ensure their cooperation with the evaluation. Whilst this recommendation may serve to improve the flow of information between involved parties in the short term and in the service of the evaluation process, in the longer term it may serve to legitimate disruptive practices and perpetuate imbalances in knowledge (Gregory , 2000).

The last of the three barriers, corresponds to a social barrier particularly to the way in which the evaluator seeks to become acquainted with the context surrounding the evaluand. . Laughlin and Broadbent (1996) suggest that, as the evaluator is the only person who has access to sufficient data to formulate a well-rounded view of the situation, this ensures his position of supremacy *vis-à-vis* other stakeholding parties. Thus, whilst Guba and Lincoln are determined that the position of the evaluator should not be privileged, based on the

methodology of Fourth Generation Evaluation this appears rather inevitable (Gregory , 2000).

3.4.4.5 Patton's Developmental Evaluation

Developmental evaluation emerged as a response to the necessity of a new approach different from the traditional summative and formative approaches for tackling situations involving social innovators working on extremely difficult problems in complex, dynamic environments (Patton, 2011). This evaluation approach brings to innovation and adaptation the processes of asking evaluative questions, applying evaluation logic, and gathering and reporting evaluative data to support project, program, initiative, product, and/or organizational development with timely feedback (Patton, 2016a). Although Developmental Evaluation does not follow a predetermined structure, for an evaluation to be considered a Developmental Evaluation needs to follow a series of eight principles specified by Patton (2016b). Among these principles, there are four which are of particular interest, the developmental, the utilization focused, the complexity, and the systems thinking principles. The developmental principle highlights, informs, and supports the pattern of development and the consequences of those patterns. The utilization focused principle corresponds exactly to the same approach of utilization focused evaluation, which is oriented towards following an evaluation focused on intended use by intended users. The complexity and the systems thinking principles correspond to a first attempt of incorporating complexity and systems thinking concepts in the design of an evaluation approach. The complexity principle directs the evaluation towards being directed by complexity premises and dynamics. Finally, the systems thinking principle, implies being attentive to interrelationships, perspectives, and boundaries of the social systems in which the social innovation is taking place and the evaluation is being conducted. This process-centric orientation to conducting evaluation contrasts the more procedural- or methodological-focused approaches that are typical of conventional evaluation (Lam & Shulha, 2014). Beyond that, developmental evaluation is methodological and method agnostic (Patton, 2011). Some examples of the use of developmental evaluation in practice, could be found in the work of Hargreaves & Podems, 2012; Dickson & Saunders, 2014; and Fagen et al., 2011. Developmental Evaluation constitutes an important step forward in the evaluation field, for two reasons, its desire to connect evaluation theory and practice, and the introduction of complexity and systems concepts not only in the practice but also on the theoretical realm of evaluation oriented towards social innovators. However, Miller (2016) poses a difficulty in doing so. It is rare to find an evaluator who can boast sufficient depth and capability in the areas Patton identifies

(by means of his eight principles) to conduct a successful Developmental Evaluation as an individual practitioner. For this reason, it is very likely that in order for Developmental Evaluation to become widely used, long time, practice, and years of study need to take place first.

3.4.4.6 Participatory, Empowerment and Collaborative Evaluation Approaches

Participatory, Empowerment, and Collaborative evaluation approaches are three different evaluation approaches which more frequently than not are confused given their similarities. According to Fetterman et al. (2017a), these three approaches have more in common than the differences that distinguish them from each other. Principles and practices, including methods, represent “common threads” uniting these approaches. In terms of principles, these approaches explicitly share many of them including stakeholder involvement, being participant focus, having capacity building as a primary aim, valuing community knowledge, and improvement. Aside of sharing principles, the three approaches use many of the same methods in practice, as they are more concerned with a systematic way of thinking and practicing evaluation than they are about the use of particular methods.

On the other hand, the main differences between the approaches are primarily related to the role of the evaluator in each approach. According to Fetterman et al. (2017b), in collaborative evaluations, evaluators are in charge of the evaluation, but they create an ongoing engagement between evaluators, and stakeholders, contributing to stronger evaluation designs, enhanced data collection and analysis, and results that stakeholders understand and see. Participative evaluations, jointly share control of the evaluation. In this type of evaluation, participants are encouraged to become involved in defining the evaluation, developing instruments, collecting and analyzing data and reporting and disseminating results (Guijt, 2014; Zukoski & Luluquisen, 2002). In empowerment evaluation, evaluators view program staff members, program participants, and community members as the ones in control of the evaluation, however they also serve as critical friend or coaches (this will be explained in the section related to empowerment evaluation) (Fetterman et al., 2017b). Each one of these approaches will be described with some detail in what follows.

Collaborative evaluation is an evaluation in which there is a substantial degree of collaboration between evaluators and stakeholders in the evaluation process, to the extent that they are willing and capable of being involved (Cousins, Donohue, & Bloom, 1996; Rodriguez-Campos & Rincones-Gomez, 2012). This approach distinguishes itself in that it

uses a sliding scale for levels of collaboration, for each different evaluation will experience different levels of collaboration. The practice of collaborate up to this moment is grounded on the use of two frameworks grounded in the American Evaluation Association Guiding Principles, these are: Collaborative evaluation techniques (O'Sullivan , 2004), and the Model for Collaborative evaluation (Rodriguez-Campos & Rincones-Gomez , 2012). The collaborative evaluation techniques proposed by O'Sullivan, are a set of techniques that have worked well with multi-site and multi-year evaluations. These techniques are summed up in a model following 4 steps in cyclical way: reviewing the program status, developing evaluation plans, implementing the evaluation, and share evaluation findings. These techniques unite perspectives of program stakeholders, evaluators and network partners to strengthen individual program evaluation strategies.

On the other hand, the second way in which the practice of collaborative evaluation could be grounded, is on the Model for Collaborative Evaluation (MCE). This model has a systematic structure and which revolves around a set of 6 components specific to conducting collaborative evaluations, providing a basis for decision making (Rodriguez-Campos, 2017). These components correspond to the identification of the situation, clarifying expectations, establishing a collective commitment, ensure open communication, encourage effective practices, and follow specific guidelines. However, they will not be described in detail in this document. According to Rodriguez-Campos & Rincones-Gomez (2012), although the MCE may give an impression to be conducted in a sequential process, it is a system that incorporates continuous feedback for redefinition and improvement in which changes in one element affect changes in other parts of the model. Some examples of the practice of collaborative evaluation could be found in the work of Fluhr et al., 2004; Rodriguez-Campos, Rincones-Gomez & Roker, 2017; O'Sullivan, 2017.

Participatory evaluation is a stakeholder-driven approach in which program planners and participants actively engage with an evaluator to implement an evaluative process. Participatory evaluation enhances evaluation use by increasing the depth and range of participation. It builds on Participatory Action Research (PAR) and participatory research models (Zukoski & Bosserman , 2017). The issue in participatory evaluation is not about which methods to use, but whose voices to include, how to include them, and determining who will speak for whom (Greene, 2000). Participatory approaches to evaluation can be further differentiated from other evaluation approaches and from each other by considering the level and extent of stakeholder involvement, the diversity among stakeholders and the

level of evaluator control over the process (Cousins & Whitmore , 1998). Decision making and leadership of the process begins with the evaluator and over time is diverted to program stakeholders (Cousins, Whitmore & Shulha, 2013; Fetterman et al., 2014). Levels of stakeholder participation vary widely from nominal to transformative, depending on the purpose of the evaluation (Cornwall, 2008; Guijt, 2014). There are many approaches to conducting evaluation. However, a 5 step model for doing so is based in the following steps: decide to use a participatory evaluation process, select and prepare an evaluation team, collaborate on making a plan, conduct data collection and analysis, and share results and develop an action plan. Stakeholders may contribute in each one of these steps. Some examples of the use of participatory evaluation could be found in the work of Zukoski, Bosserman & Baechler, 2017.

Empowerment evaluation corresponds to the use of evaluation concepts, techniques, and findings to foster improvement and self-determination by using qualitative and quantitative methodologies (Fetterman, 1996 p.4). It is an approach that "aims to increase the likelihood that programs will achieve results by increasing the capacity of program stakeholders to plan, implement, and evaluate their own programs" (Wandersman & Snell-Johns, 2005). The role of the evaluation in this type of evaluation is that of a critical friend. A critical friend, is someone who facilitates the process and steps of empowerment evaluation and who is not in charge itself of the evaluative process. Instead, the people they work with are truly in charge of the direction and execution of the evaluation (Fetterman & Wandersman , 2017). Their real role is not only to work as an advisor in the form of a critical friend but also to keep the evaluation systematic, rigorous, and on track as much as possible. However, the long term goal of Empowerment evaluation is for an organization or program to be able to evaluate its strategies on its own without the assistance of the empowerment evaluator (Cox, Keener, Woodrard, & Wandersman , 2009).

As it happens with Developmental Evaluation, Empowerment evaluation does not dictate specific steps to be followed. However, there are two models that can inform how organization and evaluators conduct empowerment evaluations, these are the 3-Step Approach (Fetterman, 2015), and the 10-Step Getting to Outcomes (Chinman , Imm, & Wandersman , 2004). The 3-Step Approach includes defining the mission, taking stock, and planning for the future. On the other hand, the 10-Step approach, which is oriented towards planning, implementing, evaluating, and sustaining strategies has the following steps: conducting a needs and resource assessment; identify the goals, population and desired

outcomes; review the literature and similar concerning best practices; determine how well the program or initiative fits the needs of the population and community; assessing the organization's capacity to implement the program; plan to implement the program; ask if the program is being implemented with quality and as planned; think about how the program is achieving its desired goals and outcomes; plan for continuous quality improvement; and review sustainability considerations. Aside of these couple of models which could be followed, Fetterman & Wandersman (2005) define a set of 10 principles to be followed by every empowerment evaluation. Some examples of Empowerment Evaluation in practice could be found in the work of Dugan, 2017; Fetterman & Ravitz, 2017; and Schnoes, Murphy-Berman & Chamber, 2000.

3.5 Final considerations

Throughout this chapter I have described different ways in which evaluation approaches have been mapped as well as the most important considerations regarding prominent evaluation approaches that have been developed so far. Taking into account the work of Guba & Lincoln (1989), Alkin & Christie (2004, 2012), and in less extent the work of Mertens (2015), I also proposed my own way of mapping such approaches by following a “waved” classification. The reason for choosing a waved interpretation of these developments was based on two elements. The first was a recognition of particular paradigms followed by different approaches, leading to four recognized paradigms which could be followed. At the same time, each paradigm was related to the degree of participation stakeholders got in the evaluation stages, going from none participation to a full participation to every stage of it. Second, using waves guaranteed a concordance between the way in which the evaluation field and the systems thinking fields were presented in this document. However, it is important to note that this only constitutes a different way of mapping approaches, which is not necessarily good or bad, but it is my own. It is important to note as well that although several evaluation approaches were covered in this chapter, there are many other approaches that could have been covered as well. These approaches were chosen because they are the most commonly used in practice, and because the work of the authors who developed them have set the foundation for a theoretical basis for the evaluation field.

Given that I have already described the systems thinking field as well as the evaluation field, I should now explore what has been done in using systems thinking concepts and methods in the practice of evaluation. The next chapter will address this issue.

Chapter 4 – Systems Thinking in Evaluation

The previous chapter described the developments that have taken place throughout the waves of systems thinking. In this chapter, we will describe how several of those approaches have been used in the evaluation field focusing primarily in program evaluation. In order to do so, we will follow the same order that was used in the preceding chapter by using the idea of waves. Following a “waved” structure might make it easier for the reader to understand the use of systems approaches that belong to each of the waves. However, it is important to note that the methodological use of these developments does not have a temporary character as the one depicted in the development of systems thinking ideas. Instead, approaches from the three waves are used indistinctly according to the needs of the evaluators and the evaluation.

4.1 Evaluation, systems thinking and complexity science. How are they related?

Throughout the evaluation literature, several types of problem situations are recognized, these are; simple, complicated and complex (Glouberman & Zimmerman, 2002; Snyder, 2013; Rogers, 2008; Patton, 2011). Williams & van’t Hof (2014) use the term “wicked situations”

to describe a complex situation as well. The complex/wicked situations are the ones of particular interest for us. However, there is not a unified definition of what a complex situation is, instead several authors have pointed out general characteristics that this type of situations might have.

According to Patton (2010), a complex situation is defined as a “situations in which how to achieve desired results is not known (high uncertainty), key stakeholders disagree about what to do and how to do it, and many factors are interacting in a dynamic environment that undermine efforts at control, making predictions and static models problematic”. Hummelbrunner (2011) & Snyder (2013) express complexity (when referring to a complex situation) as being the result of many different elements (e.g. actors, actions, factors) and their linkages and is because of these linkages that changes in a single element do not remain isolated but can influence others – often with unforeseeable consequences. In terms of Byrne & Callaghan (2014), complexity concepts include nonlinearity (small actions can produce large reactions), emergence (patterns emerge from self-organization among interacting agents), and dynamic adaptations (interacting elements and agents respond and adapt to each other). This goes in hand with Westhorp (2012, 2013) characterization of a complex situation, where he recognizes the need to understand it in terms of multiple factors in causation, non-linearity of change and emergent outcomes (both intended and unintended). Wolf-Branigin (2013) adds to these elements the fact that complex situations are self-organized. Summing up the views of the previous authors, the main characteristics of a complex situation are feedback, emergence, non-linearity in causation and adaptation, and self-organization.

If the world is complex the theories and methods of evaluation should mirror that complexity (Forss, Marra, & Schwartz, 2011) and is for this reason that systems thinking and complexity science are two ways in which complex situations in the evaluation field are started to be tackled. However, although there are several evaluators who claim to be using systems thinking and complexity (independently) ideas and methods in evaluation, the extent and ways in which evaluators are drawing on these ideas is not well understood or operationalized (Gates, 2016; Reynolds et al, 2012; Walton, 2016) . To date, there has been no systematic examination of how these ideas and methods contribute to the evaluation field or a framework for understanding when and why to use them (Gates, 2016) as the examinations of systems thinking in evaluation have focused on the concept of systems thinking and exploring the use of systems concepts and methodological approaches in specific cases in

evaluation such as the work of Williams & Imam (2006), and Williams & Hummelbrunner (2011). However, a deeper work has been conducted regarding complexity theory and evaluation. Although we will mention some of the work on this area, we will primarily focus on what has been done in relating systems thinking and the evaluation field. The reason for doing this is that the main focus of the current research work is directed towards systems thinking and not complexity, which is why exploring complexity in a deeper way might be too time consuming as it is not our central purpose.

Even though the connection of systems thinking and evaluation in practice has not been deeply explored, benefits of using systems thinking and/or complexity science in evaluation have been recognized. Both types of approaches bring benefits to the evaluation field as they share a mutual concern for (Reynolds *et al.*, 2012) challenging narrow minded reductionist practices and linearly assumed causal relationships, encouraging a dynamic more holistic perspective (however this sense of wholeness evolves throughout the systems thinking waves), stressing co-evolution for evaluation and for practices being evaluated, emphasizing the value question and seeking answers to what constitutes merit and worth and other benefits.

According to Imam, LaGoy & Williams (2006), Reynolds (2007), Hummelbrunner (2011), Reynolds *et al* (2012), and Williams & van't Hof (2014) there are three concepts of systems thinking evaluators need to know, these are: perspectives, boundaries and interrelationships. Exploring different perspectives gives an evaluator the opportunity of contemplating different ways of framing a situation. Cabrera, Colosi & Lobdell (2008) also consider the system concept as being relevant as in their own terms "any collection of related concepts can naturally be viewed as a system" (Cabrera, Colosi & Lobdell, 2008: 6). For this reason, I will address the consideration of these concepts by focusing on the four patterns presented by Cabrera, Colosi & Lobdell (2008).

Cabrera, Colosi & Lobdell (2008) present four fundamental patterns that connect the systems universe which are useful for the practice of evaluation. These patterns are; distinctions, relationships, perspectives and the system concept itself. Each of these patterns relates to the others in a particular kind of way so that none of them can be considered independently from the other three. Distinctions refer to the existence of concepts that could be differentiated, in that sense in order to make a distinction one must establish an identity and exclude other. Distinction relates to what Imam, LaGoy & Williams (2006) denote as boundaries as the

process of making distinctions involves a boundary that differentiates between what/who is in and what is outside the set boundary that is between internalities and externalities. Although a distinction is made between two concepts, there still exists a relationship between them, this brings to the fore the consideration of the second pattern, that of relationships. In Cabrera, Colosi & Lobdell (2008) terms, if there is any distinction between two concepts or agents, there must be some concept of relationship between them, at least the relation of being distinct from one another. Relationships have as a direct implication the appearance of affecting an effecting elements, affect refers to the action taken by an agent and effect refers to the result of that action on or to another entity. The sole existence of relationships between different concepts, increases the interconnectedness among the elements of a system so that its internal behavior gets to be reinforced. In that sense, once again a key pattern emerges again, in the form of system, as the simple form of system is one considered as a whole made up of at least two related parts. By having distinctions and the emergent relationships between them, a system whether it is large or not emerges as well. When talking about a system, one does not refer solely to the relationships but to the consideration of the distinctions as well. In the system pattern, the most important element is the relevance given not only to the whole but to its constituent parts as well. Finally, the perspective pattern emerges, this pattern is implicit in the consideration of a concept as it refers to a particular way of framing it. Perspective does not only refers to a particular view of conceiving a concept, but by aside of having a unique identity each of these concepts takes a point of view on its environment as well. According to Cabrera, Colosi & Lobdell (2008), the point of view of the environment upon which perspectives lie, has a human “root” perspective which includes a subject and an object relationship between its elements (concepts). The subjective character refers to the fact that any one concept (subject) cannot literally “see” another point of view, instead it interprets and attributes a particular perspective to the other (object). Perspectives relate in a particular way to the other three patterns considered by Cabrera, Colosi & Lobdell (2008) as it has the potential instantly transform whole systems, rearrange distinctions, and cause relationships to appear or disappear.

As the reader might be able to see, these four patterns are particularly relevant and helpful not only for the evaluation field but to other fields as well, as the consideration of systems thinking and its concepts represent a different way in which situations are seen. However, once the authors published their work, several other authors developed a series of reviews regarding their work. Reviews of this paper come from both, the systems thinking and the

evaluation sphere and most of them surround the following topics: the proposition of the authors of leaving aside a methodological pluralism, an attempt of unifying the systems thinking field, the practical usefulness of the DSRP rules and the possibility of achieving transformative results, the effect of bringing systems thinking conceptualization by means of the DSRP rules to the practice of evaluation, and the definition of what could be considered as systems thinking. Each of them will be reviewed in what follows.

In Nowell's (2008) and Midgley's (2008) view, Cabrera et al.'s (2008) work poses a significant challenge to the systems community by dismissing the practice of methodological pluralism and substituting it by the use of a theory of conceptualization to reconstruct the systems thinking enterprise. According to Midgley, this conceptualization represented by Distinctions, Systems, Relationships and Perspectives (DSRP) represents an attempt of unification of the systems thinking field given that they are based on recognized patterns throughout it. However, Midgley (2008) puts in doubt this unification attempt as he argues DSRP is very similar to several theories that have been used to underpin the methodological pluralism in two respects; by being multi-faceted, as it offers four concepts involved in systemic thought, and by involving a degree of "shoehorning", that is a customizing the purposes that the systems approaches were originally designed for so that they fit the given theory. Taking this into account, DSRP constitutes an interpretation imposed on other perspectives, and they are prepared to dismiss concepts in those perspectives that do not fit. This brings to the fore the same critique that was once directed to the work of authors such as Flood & Jackson (2001) when they attempted to align systems methodologies with Habermas' (1972) interests. This attempt implied a shoehorning process as well. However Midgley's main concern was that the work of Cabrera et al. (2008) limits itself to only presenting the four concepts through the DSRP without aligning any systems approaches with them. Midgley suggests that instead of opposing DSRP theory to methodological pluralism, the authors might be better off further developing their work so that they can offer a plurality of systems methods that can extend the practice of DSRP. Midgley (2008) also argues that the DSRP proposition could be considered only as half a pluralist perspective as the authors offer a multi-faceted theory that lacks of tools and techniques. This means that although it presents the four concepts by means of the DSRP it does not go further in developing ways in which they could be put into practice. This poses a challenge to the use of the DSRP.

On the other hand, there are opposed views when referring to their attempt at unifying the systems thinking field. For Datta (2008), this unification attempt is quite novel and worthy

and should be extended. However, in Midgley's (2008) view, while seemingly fine in theory, unifications are actually problematic in practice because of the limits to what any single systems thinkers can do: it is hard for any individual to work competently in the subject domain of more than a couple of disciplines. This means that when applying the DSRP rules, to any particular domain, the authors may fail in considering a wide variety of elements that may be important in practice.

Datta (2008), Rogers (2008), Wasserman (2008) and Hummelbrunner (2008) have an interest in the possibility of achieving transformative results by using the DSRP rules as well as operationalizing them. Datta, Wasserman and Rogers consider that although the authors present a strong theoretical case for the DSRP rules, they do not present any convincing practical example of their usefulness in practice and even more, of something that can convince them of truly achieving transformative results. Rogers (2008) in particular, would like to find examples of evaluation conducted in a "non-systemic" manner so that they could be compared with evaluations conducted following a systemic approach, in order to identify differences and similarities between both of them. For this reason, Wasserman proposes that by incorporating the theoretical component, developed by Cabrera et al., to the applied methodologies as well as having more practical examples of their use, will be helpful to demonstrate how the rules might be systematically operationalized, the potential benefits to be reaped, and to invite similar application to other aspects of evaluation. On this matter, Hummelbrunner (2008) presents a systemic framework based on a logic model he often uses to evaluate local and regional development programs as a tool to incorporating the DSRP rules in the practice of evaluation. This is an example of a framework which is already present in the evaluation field and which could be used for bringing to the fore element from a different field such as systems thinking.

Taking into account the need for operationalizing the DSRP rules, Reynolds (2008), Datta (2008), Nowell (2008) and Forrest (2008) find the introduction of this concepts for the practice of evaluation as being very helpful. One of the positive traits they identify of the systems concepts, is that they are simplified in four rules, which means that they are not extensive or difficult to put in practice and the fact that learning from systems thinking should not require too much time for them. As Forrest puts it, systems thinking could be seen as a "Swiss Army knife" in that sense, no one needs to master all the tools before they can benefit from having them in their toolbox. This is an important statement for systems thinking newcomers, such as evaluators, as the wide variety of concepts and the multiplicity

in which they are used, may make it more difficult for them to be familiarized with systems thinking and how to use it. According to Cabrera et al. the systems concepts incorporated in the DSRP rules are already implicit in many daily life activities not only in the evaluation field but in others as well. For this reason, in order for evaluators to practice or to incorporate systems thinking concepts in their work they will not need to learn an extensive amount of tools, methods or theory but only to use them in a conscious and regular way. Datta (2008) for example, finds some similarities in the work of Patton (2006) and Williams (2006) regarding these concepts, which is why she refers as them as being “fairly well-know”. For Forrest instead, this represents an opportunity for integrating his systems background into the activities of evaluation. The authors agree on the benefits of including systems thinking concepts on the practice of evaluation particularly enhancing the boundary setting processes that take place in evaluation, and thinking and communicating with regard to complex issues which are continually tackled in evaluation. However, as I stated earlier in this section, there is a need of extending the practical examples of their use, so that the DSRP rules are really seen as beneficial. For Forrest (2008) the role of systems thinking in evaluation is particularly important if the latter has a shift in its orientation moving on from a historically oriented base towards a future oriented one, particularly because systems thinking helps to understand which might be drivers of change and for making inferences on how things may work in the future. A shift in orientation may give systems thinking a more important role in evaluation. Finally, Reynolds (2008) considers the work of Cabrera et al. (2008) as enriching in terms of our understanding of what is considered as an affect or an effect among system parts and between a system and a wider system. This is conceptually recognized as enriching particularly because Reynolds (2008) perceives his own experience as being limited by being predominantly focused on dominant causal interrelationships and a focus on factors affecting the system as against ethically important effects of a system on its “victims” (Reynolds, 2004; 2007).

Boundaries are a key element in systems thinking approaches more explicitly in those approaches in the third wave of systems thinking, as the establishment of a particular boundary drives how we see a system as it determines what and who is being included and excluded. The importance of reflecting upon boundaries lies in the necessity of assessing the consequences of the boundary choices (boundary critique) (Boyd et al., 2007; Ulrich, 1988). The establishment of boundaries is fundamentally about values as they are judgments of worth. This goes in hand with the basic definition of evaluation as it is defined as a

systematic process to determine merit, worth, value or significance (American Evaluation Association, 2014). Here, it is important to take into account that the relation between the systemic and the systematic concepts need to be addressed as they are opposed in definition and sometimes their joint use might seem contradictory. Third is the concept of interrelationships or entangled systems referring to the existence of a system within a bigger system and the fact that a system cannot be analysed on its own. Instead it needs to be considered in relation to other systems, whether these are wider or within it. In considering this interrelationships, a systems thinker is able to critique the current boundaries of the systems and decide if they should be widen, kept or narrowed.

The systems thinking characteristics that we just described give us the possibility of exploring how they can be identified when using approaches of the different systems thinking waves. Authors such as Reynolds (2007) consider that each wave of systems thinking is particularly focused on one of these characteristics. According to Reynolds, the approaches in the first wave of systems thinking, are primarily focused on understanding the complex interrelationships that might be present in a system, the second wave approaches are focused on including multiple perspectives and the third wave approaches are related to the critique of different boundaries of the system.

We will argue that the approaches provided by the third wave of systems thinking are the ones that provide the more complete insight in conducting program evaluation, as these approaches, particularly boundary critique, when used properly can not only provide an insight of what and who should be included in an evaluation, but also why an evaluation was shaped and designed in a particular way. These approaches are also useful for reflecting on the multi-methodological character of interventions, particularly on the appropriate use of methods in order to serve specific interests and purposes and attempting to keep a theoretical coherence among them (Mingers & Brocklesby , 1995).

We will now start how the approaches of each wave were used in conducting evaluation. Two types of interconnection between the two fields are described below. The first one is depicted on practical examples on how evaluators used systems thinking methods to conduct some sort of evaluation (total or partial). Some examples of these examples might be more descriptive than others mainly due to the amount of detail given by the authors in the publications that were revised. The second type of interconnection takes place when systems thinking or a particular approach of systems thinking is used to expand knowledge through the generation

of frameworks and methodologies for conducting evaluation. Some of these approaches are presented as well.

4.2 First wave of Systems Thinking and its application on evaluation

Different approaches from the first wave of systems thinking are commonly used to conduct an evaluation. Taking into account the description of the approaches given in the previous chapter, the approaches within the first wave of systems thinking that are commonly used in evaluation are; Systems Dynamics, the Viable Systems Model, and complexity science approaches. Although complexity science can be seen as parallel to systems thinking, we approach it as being part of the first wave of systems thinking.

4.2.1 Systems Dynamics and Evaluation

Authors such as Burke (2006), Williams & Hummelbrunner (2011), and Newman *et al* (2003) present examples of the use of Systems Dynamics (SD) to conduct program evaluation. Burke uses SD in the planning stage of a teacher's professional development program where he wants to understand how the interaction between the program components affects the ability of the program to build teachers capacities within a particular period of time, and also on which are the costs associated with the form of teaching promoted by the program. In order to develop the SD simulations, the author relies on available quantitative values for the important variables in the simulation and estimates derived for the qualitative variables. Burke (2006) argues that the need of using SD lies on the fact that other available tools to identify and evaluate the strengths and weaknesses of critical factors in multicomponent programs are not useful to understand in what ways the factors interact or to integrate the functioning of the separate components of the program. Burke recognizes as major strengths of using SD its ability to capture nonlinear dynamics, identify unanticipated consequence of system behaviour, and rapidly test policy options and resource allocations. The use of SD gave him good results in understanding the program's structure and its effectiveness.

On the other hand, Williams & Hummelbrunner (2011) present the use of SD in a micro loan scheme that was part of a HIV program of a Non-Governmental Organization (NGO). In this case, the interests was to understand the reasons of the variations in popularity of the scheme presented to the beneficiaries and what strategies could be adopted to reduce that behavior. Later on in this chapter we will describe how the program as a whole was evaluated through the use of Critical Systems Heuristics. In order to use the SD approach, evaluators first used

Causal Loop Diagrams (CSD) as this was easier to use with the stakeholders that were being included in the discussions for representing the situation. After the CSD were developed with the help of the stakeholders, the evaluators proceeded to simulate the diagrams by converting them into a flow stock model of SD. In the words of Williams & Hummelbrunner (2011), in order to ensure the model reflected the behavior seen in the real world, they “played around with it to see what insights it gave about the issue”, at this point the stakeholders of the program were no longer participating on the simulation process. The use of the model did give them an insight of where the solution for the fluctuation might have been, but it did not give them an idea of how or what the NGO could do to solve them. On the use of SD, the authors recognize that the insights from SD models can be very valuable for evaluators and policy makers alike. For instance, for understanding that elements of a system interact through feedback loops, that causes and effects are not often linked directly in time, and that delays can cause unanticipated behaviour. Just as Burke (2006), they recognize that the ability to capture nonlinear dynamics, identify unanticipated consequence of system behaviour, and rapidly test policy options and resource allocations can make SD an important tool for many evaluation studies.

However, Williams & Hummelbrunner recognize some pitfalls as well, being the most important one regarding SD models as representations of reality rather than as aids for the development of deeper understanding. Taking this representation for granted is a problem for several reasons. First, a model does not have an adaptive character, which implies that conceiving a model as a representation of reality gives a static character to the relationships depicted through it, being incapable of adapting to changes in the environment that might affect what is being represented through the model. Second, when taking for granted that models are representations of reality, it might be assumed that the way it depicts relationships might be infallible meaning that a different representation of reality is not possible. However, models can be very useful to deepen understanding of a situation when properly used. If an infallible character is not given to a model, it can be particularly useful to understand a stakeholder perspective (the one producing the model) regarding the relationships taking place in the situation. Using more than a single model can help the different stakeholders to understand not only how a situation might be unfolding but also how a situation is being seen from the eyes of other stakeholders.

Another major concern for Williams & Hummelbrunner is that the modelling process and the insights generated tend to be expert’s work, leading to neglecting the involvement of other

stakeholders. This concern is directly related to their own work, as stakeholders were only included in the initial stages when producing CLD, and this was changed in further stages as the evaluators proceed to use SD models. However, in the development of SD theory authors such as Sterman (2000) set as one of the main features of the development of SD the involvement of clients as early and deeply as possible, arguing that “models built out of the sight of the client will never lead to change in deeply held mental models and therefore will not change client behaviour...Encourage them to suggest and run their own tests and to criticize the model (Sterman, 2000:80)”. For this reason, one could say that the concern is more related to the way they use the approach and not to the way it has been theoretically developed in recent years.

Additionally, Newman *et al* (2003) use SD to monitor the progress of achieving targets in the implementation of a malaria control program by arguing that a more rapid progress can be achieved through the integration of different activities and the explicit link of several actions. The authors recognize that a well-designed model can generate graphical projections of the impact of the proposed policy action on key indicators and target results permit an estimate of financing requirements in order to achieve those results, and serve as a useful tool for a constructive dialogue on prioritizing policy interventions. However, they also address several consideration and caveats when using a model based approach such as SD. First, they recognize that there are multiple ways of reaching a goal and achieving a particular target, which means that the relationships depicted by a model only represent a particular way of viewing and assessing a situation. In their own words, “there may be many possible paths, but with a public policy focus, what is important is modelling and analysing the path chosen by public agencies and authorities” (Newman et al, 2003: 13).

Second, the simulations that can be run based on the model developed are only valid for analysing small changes around a local optimum. This means that if an evaluator is considering wider changes for a particular situation, a simulation may not capture them, leading the model as a useless approach to the situation. Third, the authors also consider that when using SD, constant comparison of what one thinks is going to happen with what actually happened is required so that a constant modification of the approach is generated otherwise the model will become victim of its capacity to adapt to real life changes taking place in a particular time. Fourth, it may be difficult to model what policy actions lead to results. This is primarily related to the way a situation is being understood and the problem situation is being considered because if the mental model of the situation is not clear, one

might not be able to discuss, analyse and debate about it. Other examples of the use of systems dynamics in program evaluation are Frederick, Deegan & Carman (2008), Martin & MacDonald (2012), and Derwisch & Lowe (2015).

As can be seen through the general description of the previous examples, the use of SD in program evaluation is not limited to a specific area of knowledge or practice, instead it can be used in a wide variety of field and for different purposes. All of the authors highlight the importance of using SD as the situations they are facing have a multidisciplinary and multilinear character which cannot be easily understood without depicting it with a model. Using models and simulations can be useful to predict the future behavior of a program or situation taking into account its original design, also allows testing different hypothesis of causality of the impact of policy options on outcomes which allows the evaluation team to consider multiple alternatives on how to proceed in the planning and implementation of a program. However, it becomes useless if the evaluator does not adapt the model to further real life changes which is something very likely to happen.

Even though the different authors recognized some benefits of using SD they also repetitively recognized some shortcoming that support our proposal of approaching program evaluation from a third wave perspective, implying a critical approach. First and most important of all, when using SD the involvement of multiple stakeholders in the different phases of the evaluation might be limited not only by the willingness of the evaluator and those who commission the evaluation to include stakeholders that are being affected by the program but also because the knowledge required to be included in the different stages of the evaluation might supersede the knowledge of the stakeholders. In this way, SD might work as a tool to foster the power of those who command the evaluation and hinder the involvement of those who are directly affected by it. Using models requires some type of preparation or expertise in the field, not only to be able to produce them by using simulation programs but also to be able to interpret the results of the simulations, graphs etc. In practical terms, if the models are to be applied and used with other stakeholders, the power to interpret and make decisions over a particular state of a situation might rely entirely on the ones who have the adequate knowledge, which are very likely to be the evaluators or the ones who commanded the evaluation. This knowledge gap represents a disadvantage not only for the interpretation of the results but also for the production of the model which ultimately has an impact on the quality of the analysis and outcomes expected from the evaluation that is being conducted. If the evaluator is producing the model on his own, it is very likely that he will miss real life

connection and interactions that can only be seen through the eyes of those who are directly involved in the program and who have been part of it for longer time.

Second, although models might be useful to provide an insight of the situation, they cannot be regarded as representations of reality, they can only be conceived as an aid for the depiction of reality. In this sense, these types of models can help to understand the general condition of a situation but they cannot depict qualitative aspects of it, as for example which are the interests, goals and perceptions of the human beings that are being modelled, and in many occasions key relations that are explicitly considered by a model might be affecting the performance of the program but the evaluators may not be able to see them as they do not explore the qualitative character of the situation. Here, it is important to consider the fact that SD might fall into the mistake of objectifying human beings while attempting to include them as parts of a simulation (Jackson, 1982, 1987). And as was shown in the examples, the qualitative variables that might be considered in a model are approaches from a quantitative view. Here again, the depiction of reality is being completely altered by trying to give a numerical approximation to something that cannot be directly quantifiable.

As a particular concern we highlight the statement presented in the work of Newman et al (2006) as the authors recognize the view of the policy makers as the only one valid to be depicted in the development of a SD model. This statement is a clear example of an inadequate management of a program that involves the community participation as this means that their real needs might not be considered and even worst, a program might be being considered successful only on the basis of the results of a model and by the quantitative results that it might be giving.

4.2.2 The Viable System Model and Evaluation

While exploring the use of the Viable Systems Model (VSM) in evaluation, we find examples in the work of Fitch (2006), Hart & Paucar-Caceres (2017), and Tepe & Haslett (2002). It is important to clarify that most of the work of the past fifty years regarding VSM has been directed towards evaluation, however VSM literature has not been widely connected to evaluation literature. Further examples of the work in VSM can be found in the work of Leonard (2009), Adham et al., (2011), Sharma & Gutierrez (2010), Morell (1984), and Haslett & Sarah (2006). A particular characteristic of these approaches is that all of them use the VSM to conduct evaluations related to the use of some sort of management information systems, this might be related to the fact that the VSM is particularly interested with the

communication channels and the information flow in organizations. In the case of Fitch (2006), he examines how a non-profit organization system of information supports the organization's ability to process information, and in order to produce the 5 systems of the VSM, he uses as sources of information interviews, review of case forms and paperwork, observation of the staff in their work environment, informal conversations with clients, and develops a semi-structured questionnaire applied to relevant stakeholders, although he does not specify how these stakeholders had been selected. Fitch recognizes that a "cybernetic evaluation" can be performed as prior to designing and installing a management information system or after it is already in place to determine what specific aspects of its design need to be modified for efficiency.

On the other hand, a particular characteristic of the work of Hart & Paucar-Caceres (2017) is that they chose a particular evaluation approach to conduct their evaluation, this does not happen with any other of the authors as they only specify the type of evaluation being conducted in terms of it being formative or summative. Hart & Paucar-Caceres, use the VSM and a Utilization-Focused Evaluation approach to support a participative theory of change directed towards evaluating a technology supported learning program. As they combine an evaluation and a systems approach, the authors highlight the importance of the systems concepts described by Imam, LaGoy & Williams (2006), Reynolds (2007) and Hummelbrunner (2011) presented at the beginning of this chapter for categorizing an inquiry as being systemic. One of these three elements is setting boundaries, as being critical about boundaries will determine what and who are being included. However, when using a Utilization-Focused approach, the reflection on boundaries is conditioned to the interests and perceptions of the users of the evaluation. As this type of evaluation implies that the evaluation users are the ones that determine the scope of interest relevant to their purpose and needs, and the role of the evaluator is presenting a "menu" that fulfills established evaluation standards and principles (Patton, 1986, 2012). For this reason, the approach uses a logic model to reflect on these boundaries. However, the authors themselves recognize that the type of logic model that was used has as a major weakness referring to the possibility of it being too simplistic in representing relationships between components as linear. It is also important to note that the initial recognition of the possible participants of the evaluative process by the evaluation facilitator is conducted by using a "snowball approach" and is suggested to be guided by conceptualizing where the various stakeholders may fit in undertaking these VSM functions, their role in decision making, and hence their potential

interest in the evaluation.

The authors were able to recognize the importance of several of the key concepts of VSM in the situation being evaluated. For example, the prior knowledge, motivation and experiences of the evaluation facilitators influencing their interpretation and behavior was seen as a source of variety of the system. According to Hart & Paucar-Caceres (2017), concepts derived from VSM have helped to draw out stakeholders' articulation of a theory of change which in turn has given some clarity about the level of granularity (system in focus), represent their perceptions of its coupling arrangement with the wider environment and higher level of recursion, focus on important variety generators and strategies for managing variety that will influence what the system does, and reflect their perceptions of who the other stakeholders are and their interest.

On the other hand, Tepe & Haslett (2002) explore the situation of an organization that was facing several issues related to the underutilization of a computer reporting system, the workers compensations costs and the capacity of the organization to show responsibility towards accountability. By using VSM the evaluators realized that there was little concordance between reality and what a VSM would expect. Although VSM was important in highlighting who should produce the policy, who must receive the feedback and how important intelligence about the environment is, it was not useful to sort out the problems the researchers encountered. In order to address them, the authors appeal to the use of a variety of approaches such as Soft Systems Methodology, Balance Scorecard and others. However, there is no further detail on how these approaches were employed. The authors recognize that there were historical, managerial, motivational, and cultural reasons that prevented the open, honest, and accurate provision of information about safety issues (Tepe and Haslett 2002) but they do not go deeper on describing how this took place, and how it affected them.

When exploring the previous examples, it can be said that in general terms, VSM is useful to conduct evaluation because it focuses on both the structure and process of information flow through organizations and it can be used as a diagnostic tool as well as a design tool. However, there are problems measuring how well a particular variable is actually performing as subjective judgment from a decision-maker is involved (Jackson, 1988: 9). This highlights the fact that much of the apparent power of the elaborate information and control systems, detailed in the VSM, dissolves when the difficulties of matching the model's demands to real-world organizations are recognized (Jackson, 1988: 10). Moreover, the approach has

difficulties handling the existence of multiple perspectives which means that an adequate provision of information might be limited by inadequately addressing issues related to multiple perspectives such as values, cultural characteristics, interests etc. Checkland (1980) puts it as the VSM taking the organizations as following a machine type. This representation misses the essential character of organizations - the fact that their component parts are human beings who can attribute meaning to their situation and can therefore see in organizations whatever purposes they wish and make of organizations whatever they will. Because of this, it is as legitimate to regard an organization as a social grouping, an appreciative system or a power struggle as it is to see it as a machine (Checkland, 1980, 1986).

According to Ulrich (1981, 1983) Beer conceives the design of VSM as a tool design and not as a social systems design, lending itself to an autocratic usage. Ulrich argues that while Beer's model is capable of generating a degree of 'intrinsic control' - spreading the sources of control throughout the architecture of the system, it cannot generate 'intrinsic motivation' - distributing the source determining the system's goal-state and purpose throughout the system. Only the top controller is in a position to change the system's basic orientation according to its own particular desires. For Ulrich, the model provides tools which are capable of being misused and, in all likelihood, will therefore be misused (Jackson, 1988: 13). Ulrich (1981) also believes that the VSM is unable to facilitate debate about the nature of the goals pursued. According to Jackson (1988), in Beer's view, good management can be no more than management that establishes requisite variety between itself and the operations managed, and between the organization as a whole and its environment.

As can be seen through the preceding examples, the approaches belonging to the first wave of systems thinking are used primarily with the aim of comprehending interrelationships, just as was stated by Reynolds (2007). Although these approaches might be very useful to make a situation clearer, in our opinion they lack the sufficient richness to conduct the evaluation of a social program. In general terms, programs and organizations imply the constant interaction of different stakeholders and a dynamic environment. These elements imply that the way of tackling the situation needs to have a dynamic adaptable character as well. First wave approaches although having a degree of dynamism, in terms of considering interrelationships, are not directed towards managing the existence of multiple perspectives or being capable of adapting to the changing character of the situation. Rather, these approaches follow a black box kind of structure which takes an input (which in this case is most likely to be data) and transforms it according to the type of model that is being used producing an output. Its

adaptiveness is more related to the change in the data than to the capacity of changing the structure of the model which is being used.

On the other hand, these approaches are mainly focused on modelling, and this might not be the best way of tackling social situations. Better depth can be gained when framing a situation from more than one perspective as this multiplicity is the one that enriches the picture of the situation that is being assessed and is the one that lead the evaluation in a particular direction or another. When approaching a situation from a particular point of view, the evaluator is taking for granted all the assumptions given by that way of framing the situation and is somehow giving a power position to the stakeholder whose view is being taken for granted. Aside of these consideration, the authors who worked with the systems approaches in the preceding examples, show a deficient ability to design an inclusive evaluation approach as some of them do not even attempt to include the stakeholders in the evaluative process and others although trying to include them, they end up relegating the stakeholders.

4.2.3 Complexity Science and Evaluation

According to Gates (2016) complexity science typically takes two meanings: an alternative paradigm of science; and a set of ideas, theories, approaches and methods for studying complex systems. These are phenomena where interconnections between elements are nonlinear and the emergent properties from interconnections are subject to uncertainty (Reynolds, et al., 2012). Complexity theory conceives of systems that are open and evolving over time to create path dependence, but subject to internal perturbations which inspire nonlinearities. This means that, while prediction is not possible, explanation is (Callaghan , 2008). This type of approaches often use agent-based modeling created to simulate environments through computer intensive techniques and replicate the interactions of individuals (agents). Based on these agent-level activities and interactions with others, it becomes possible to study the layers of emergent behaviors produced by agent-level interactions (Wolf-Branigin , 2013). Complexity science is concerned with relationships and the interconnectedness of people. That being said, the work of Eoyang (2006), and Wolf-Branigin (2013) are examples of the use of complexity approaches for serving evaluation. Further account of complexity theory in evaluation can be found in the work of Sanderson (2000), Barnes (2003), Byrne (2013) and Westhorp (2012).

Eoyang (2006) depicts the use of Complex Adaptive Systems (CAS) and Human Systems Dynamics (HSD) in evaluating an organizational change process in a county social services

department and to recommend ways to move the change forward more effectively and efficiently. CAS were used to describe a behavioral pattern in the organization under study, attributing the following characteristics to it:

- Participants possess a moderate degree of freedom of action
- The system of interest is defined more in terms of functions than structures
- Change is seen as dynamic
- The system is open
- Change is acknowledgeable at multiple levels of analysis
- Expectations and outcomes are emergent rather than determined

These characteristics raise the possibility of considering the system as a CAS and in that order, HSD could be used to articulate the CAS with social sciences as CAS are traditionally used in a mathematical and physics realm. HSD provide three conditions (a container, differences within the container, and exchanges among agents) for self-organizing human systems, this triad constituted the CDE model which can provide insight of interrelationships, emergent patterns and options of action, however it cannot be used to predict or control system outcomes (Eoyang, 2006).

On the other, Wolf-Branigin (2013) presented a set of short examples where the evaluation of services developed for and with people with disabilities took place. The examples were directed towards a historical analysis of the emergence of disability policy, a complex systems approach to organizational accreditation, assessing the community needs through travel training, and analysing housing patterns of disabled people. Each one of the examples emphasized different complexity theory characteristics. In the case of the historical analysis, the main focus was on the use of feedback between the service providers and the families of those receiving the services, in order to establish the primary goals that should be assessed. The second example, used complexity concepts for understanding how an organization emerges, what attracts clients (agents), how these agents self-organize, how the system uses feedback, and how an emergent behaviour results. Further use of complexity in the example was oriented towards the use of social networks as a non-linear approach to outcomes.

The third example, studied four properties of organizations or systems: intention, boundaries, resources, and exchange, as suggested by Katz and Gartner (1988). These properties were used to facilitate information sharing among the participating organizations. However, the exploration of boundaries was limited to the physical boundaries that limited the system, for

example the city limits or the bus lines that individuals could use, not exactly to how to frame the intervention. On the other hand, Wolf-Branigin (2013: Chp 5, 12) states “we applied a systems evaluation framework focusing on the three waves of systems theory: general systems theory, cybernetics, and complexity science (Midgley, 2006)”. However, although the author does not explain how he did it, we will argue that Midgley does not consider general systems theory, cybernetics and complexity science as being the three waves of systems theory. Instead they all belong to the first wave of systems thinking and General Systems Theory constitutes one of the pioneering works for the development of systems thinking as a field. This might be seen as an opportunity for introducing systems thinking in the evaluation field in a more formal way, one that could make clear to the authors how systems ideas were developed, evolved and how they might be useful in the evaluation field. Although authors in the evaluation field do not have the obligation of knowing how systems thinking operates, it will be very useful for them to have a deeper insight on the subject so that systems concepts could gain relevance in the field.

The last example of the work of Wolf-Branigin analysed housing patterns through the use of a geographical spatial autocorrelation approach, and through a k-means clustering. Exploratory and interactive spatial analysis often relies on an inductive approach rather than deductive hypothesis-testing. This evaluation used spatial analysis to measure community inclusion.

One of the most important examples relating complexity with program evaluation theory is the work of Patton developing the Developmental evaluation approach. According to Patton (2008), Developmental evaluation refers to long-term, partnering relationships between evaluators and those engaged in innovative initiatives and development. Developmental evaluation processes include asking evaluative questions and gathering information to provide feedback and support developmental decision-making and course corrections along the emergent path. The evaluator is part of a team whose members collaborate to conceptualize, design and test new approaches in a long-term, on-going process of continuous improvement, adaptation, and intentional change. The evaluator’s primary function in the team is to elucidate team discussions with evaluative questions, data and logic, and to facilitate data-based assessments and decision-making in the unfolding and developmental processes of innovation.

At its core is the idea that evaluations cannot be conducted in a linear way, as not all

problems are bounded, have optimal solutions or occur within stable parameters (these are complex problems). This type of evaluation is directed towards supporting the process of innovation within organizations and in its activities in ways that enable exploration and development. It facilitates assessments of where things are and reveals how things are unfolding; helps to discern which directions hold promise and which ought to be abandoned; and suggests what new experiments should be tried (Gamble, 2008).

Developmental evaluation may also consider the dynamics of collaboration itself. Complex problems tend to require the integration of diverse perspectives from different parts of a system. Various stakeholders may understand the problem differently and enter into an initiative with diverse reference points. Within this diversity, there is still a need to develop and execute strategies. Developmental Evaluation helps collaborators to recognize and work through differences in perception that might otherwise fragment the work and hamper ongoing developments.

Ultimately developmental evaluation is about rigorous inquiry for development. It is being intentional about using data in a meaningful way to inform innovation in progress. The product or result of a successful developmental evaluation process is informed changes in what is being evaluated (Dozois, Langlois, & Blanchet-Cohen, 2010).

Complexity and evaluation examples are widely focused on feedback, emergence and modelling, however as it happened with the other approaches of the first wave of systems thinking, the authors recognize that it is useful to identify, characterize and describe situation but in order to be able to respond and suggest recommendations, practitioners need to use a wider set of tools. Although the examples explored above were directed towards social program evaluations, the direct participation of the stakeholders in the evaluation process (aside of being used to collect data) does not take place. An important considerations need to be made, it is the fact that in the evaluation field viewing a situation as complex tends to be interpreted as tackling the situation through the use of complexity science and this is not always the case.

4.3 Second Systems Thinking Wave approaches in evaluation

Among the approaches of the second wave of systems thinking, the use of the Soft System Methodology (SSM) is the one that is more commonly related to the evaluation field. Other common applications are related to Strategic Assumption Surfacing and Test (SAST), however we will not discuss SAST developments as we did not describe it in the previous

chapter. Some examples of the use of SSM in evaluation can be found in the work of Attemborough (2006), Williams & Hummelbrunner (2011) and Hou Tay & Kee Pong Lim (2006) which use SSM in a dialectical way. Attemborough uses this approach in a formative evaluation of the partial achievement that had been reached in a community sector where an urban regeneration scheme was being completed. Williams & Hummelbrunner also present a case involving communities, however this case differs from the first one in that the role of the evaluator was related to the need to find the most appropriate way of evaluating the performance of organizations that were being given money by the government to provide services to marginalized communities. Both cases have in common the consideration of multiple stakeholders, however they approach the construction of the rich picture and the CATWOE of the situation in a very different way. In the case of Attemborough, the author relies on reports, statistics, memos and other written tools to construct multiple rich pictures of the situation. In the case of Williams & Hummelbrunner, the authors did not rely on written information but built their rich picture with the help of the stakeholders. Another major difference between the two authors is that in the first case the author chose a particular perspective from where to see the evaluation of the program before developing a CATWOE. In the second case, the authors develop a set of holons which implied multiple perspectives from which the program could be viewed and for each of these perspectives a CATWOE was developed.

In both cases, the authors consider the application of SSM as being successful because it helped them view major differences between the ideal situation and the reality, giving them the opportunity to implement changes that would lead to improvements in the systems being considered. At a practical level, it might be a very difficult task to ask the evaluators and those in charge of commanding the evaluation to adopt a different perspective from their own. The adoption of different perspectives would be associated to the interests and values of the participants and in how convenient might be the adoption of a different perspective for those who command the evaluation and those executing it. As in the case of the application of approaches belonging to the first wave of systems thinking, this is related to how power related to the position of the stakeholders is handled. For this same reason is that the approaches belonging to the second wave of systems thinking are useful for situation contexts where coercion is not likely to be present. However, the issues of power differ in the first and the second wave of systems thinking, as in the first wave issues are related to expert dominance and in the second wave these are related to pseudo-participation, on the inability

for stakeholders to talk openly.

On the other hand, Hou Tay & Kee Pong Lim use SSM in a different way. The authors use a Dialectic SSM to conduct a self-evaluation on a set of existing practices (“the actual processes”) against a set of guidelines (“a set of ideals”) adopted by a railway provider. A dialectic view of SSM does not mean that it is different from SSM but that it is seen from a different perspective. According to Dick & Swepson (1994) SSM can be understood as a learning cycle that can be seen as progressing through four dialects, these are: between the situation and a description of the essence of such situation, namely the root definitions, between the root definitions and the conceptual model, between the conceptual model and the real situation out of which possible improvements to such situation may arise, and between plans and implementation (For further information regarding Dialectic SSM see Dick (1993), Dick & Swepson (1994), Dick (2000), Sankaran, Tay & Cheah (2003), Sankaran & Tay (2003), Tay & Lim (2004)).

In the particular context of the situation, the dialectical process took place between three members of the railway company, one of them being an expert and the two others knowing little about the subject of study. The first dialectic process took place iteratively between the three individuals with the aim of gaining clarification of the situation for the two individuals who lack the knowledge regarding it, this process took place iteratively until the other worker considered the situation was clear enough for the other individuals and among the processes that took place was the development of the CATWOE. The second dialectic process used the information of the first dialectic process to produce an ideal model of the situation, in the particular situation this referred to a step-by-step procedure for a particular operation to be carried out in the railways. The third dialectical process was focused on transforming the step-by-step procedure developed in the previous step into a program and its validation. The final dialectical process involved the dissemination and an ongoing evaluation of the program.

In the dialectical process, an attempt is made to handle the inclusion of the different perspectives of the participant, although this is not made in the evaluative process but in the creative process of the program. However, when proposing the ideal program created by using SSM, power and conflict might appear in the willingness to adapt the proposed program in the organization.

4.4 Third Systems Thinking Wave approaches in evaluation

The work regarding the use of the third wave of systems thinking in evaluation is primarily related to the Critical Systems Heuristics (CSH), and Systemic Intervention. The use of these ideas has gone further than simply using them as methodological tools, as they have also been used to develop frameworks aiming at facilitating the proper application of an evaluation.

4.4.1 Critical Systems Heuristics and Evaluation

Examples of the use of CSH can be found in the case of a NGO described by Williams & Hummelbrunner (2011), and in the work of Reynolds (2006 and 2010) using CSH in different fields. Williams & Hummelbrunner (2011) present the case of a European aid agency which was supporting a project to reduce the incidence of HIV and other sexually transmitted infections in Africa within sex workers operating from different bases (part of this case was described in the first wave section). The particular interests were directed towards the fact that these workers were knowledgeable about the diseases through a peer-educator program, that they could be able to have alternative sources of income via a micro-credit business loan scheme, and that they could have access to suitable reproductive health services and have appropriate attitudes and life skills, via improved STD health services.

The project had four indicators of performance aiming at measuring the changes produced by the project in terms of the number of sex workers who were relying less on sex for income, who had access to quality health services, who were correctly and consistently using condoms and the proportion of sex workers who were aware of modes of transmission of HIV. Although these indicators were useful to measure quantifiable changes in issues related to the number of women involved in sex work, its use led to the proposition of a series of questions about the operation of the project, primarily focused on whether it had been properly designed, planned and executed since the beginning. The evaluators used CSH to answer these questions. However, the authors do not give further detail of the customization and application of the questions.

As a result of the use of CSH, Williams & Hummelbrunner concluded that the project had initially been developed with little investigation of the local circumstances and boundaries were set according to a technocratic approach to a complex sociotechnical situation. Moreover, certain interests were marginalized by boundary decisions taken uncritically as the project excluded key stakeholders from its different areas of operation. Aside of this, according to Williams & Hummelbrunner the use of CSH added three important features.

First, it critiqued the expertise and the notion of guarantee as it questioned the rightness of those who were seen as experts. Second, it added legitimacy as the legitimation of the system could only be given by those outside of it, and third it highlighted the importance of drawing boundaries. In this particular case it could be noted that inappropriately drawing boundaries since the beginning of the project led to a low performance of it.

The work of Reynolds extensively includes the use of CSH in evaluation, here we will briefly explore two examples of his work. Reynolds (2006) describes an evaluation exercise that took place during the 1990's in three different participatory planning projects for rural development in Botswana. The application of the CSH questions was customized for each one of the projects according to the perceived stakeholder role, the particular level of planning being focused upon and the information arising from prior interviews with other stakeholders and relevant literature. Reynolds (2006) states that CSH could be used in four different modes when related to evaluation, these are; as ideal mapping (only using the "ought" form), as summative evaluation (contrasting the "ought" with the "is" to generate critique), as formative evaluation (allowing critique to inform new understandings and practices) and as probing evaluation (contrasting "ought" with "is" as a means of questioning dominant or hegemonic premises and judgements). Probing evaluation differs from summative evaluation in that the former contrasts the "ought" with "it" with a diagnostic purpose, whereas the latter uses such contrast with the purpose of generating critique. In this particular case, he approached the evaluation as a formative one.

Another example of the work of Reynolds, is his 2010 work where he used CSH and boundary critique to contribute to Michael Patton ideas on developmental evaluation, focusing primarily on stakeholder development. This particular work of Reynolds, does not represent an example of CSH per se but constitutes an application of CSH to enrich the ideas and practice of developmental evaluation. According to Reynolds (2010), a purposeful developmental evaluation from a systems thinking in practice perspective based on boundary critique must continually question boundaries, beginning with the first boundary question regarding purpose. The questioning can be regarded as an ongoing conversation that should strive to solve three fundamental questions:

- What is at stake?
- Who are the stakeholders?
- What possibilities exist for improving stakeholding?

The author suggests that the conversation should be conducted by using the CSH reference system and directed towards two forms of boundary critique; boundary reflection and boundary discourse. The former is focused on understanding the judgments made in real world situations of interest from different stakeholders perspectives, whereas the latter focuses on the more difficult processes of making judgments amongst stakeholders associated with a situation of interest. When using CSH as a reference system, the emphasis moves from one of identifying “stakeholders” in an undefined situation, towards a focus on “stakes” associated with a more clearly defined process of reflection on the system of interest.

Reynolds (2006) and Ulrich & Reynolds (2010) recognize several benefits of using CSH for conducting evaluation. First, the boundary questions try to make explicit the boundaries that circumscribe our understanding. These boundaries inform our thinking about values and motivations built into our views of situations and efforts to improve them, power structures influencing what is considered a “problem” and the knowledge basis defining what counts as relevant information. Second, the questions offer a way to examine how we frame situations. In this way, people can be aware of how they are viewing a situation and in that way it is easier to examine the multiple perspectives that people bring into a situation. Third, CSH promotes effective practice as they not only allow us to analyze a situation but they let us change them as well. Boundary questions allow us to understand which are the boundary judgments at work, so that everyone can understand the implications of them enabling a practical focus on ways of improving the situation.

The authors also recognize the importance of critique as a platform for improving our understanding and practice associated with a situation of interest. It’s importance is highlighted by the necessity of having the adequate skills in translating findings and producing reports as they foster improvement and signal point of inter-relatedness.

4.4.2 Systemic evaluation

The work on this topic is directly oriented towards the production of frameworks. The first attempt to produce a framework was carried out in the work of Gregory & Jackson (1992), where the authors designed a grid type matrix for choosing evaluation models. However, both of the authors recognized some gaps on their work, which is why in 1997 Gregory presented an extension of the initial work. The critique of other authors of their work lead to the design of a new systemic evaluation introduced in 2007 in the work of Boyd et al., which not only critiqued the previous work of Gregory & Jackson (1992), and Gregory (1997) but also

reframed it. The work of Midgley (1989), Midgley (1990), and Midgley & Floyd (1990) is also relevant. These developments will be described in the following section.

Gregory & Jackson (1992) developed a framework for categorizing evaluation approaches by using systems thinking tools. They recognized that particular evaluation approaches might be more suitable for some situation than others according to the characteristics of the evaluation party (an evaluation party is the group of people directly concerned with carrying out the evaluation, it might be represented by a single individual or by a group of individuals with different interests in the organization) generating different evaluation contexts. In order to produce their framework, the authors recognize two key variables as affecting the type of evaluation to be conducted, these are; the way the evaluation party sees organizations and the evaluation party variety.

Regarding the way the evaluation party sees organizations, the authors rely on their evaluation experience and in the sociological paradigms developed by Burrell & Morgan (1979) to classify the possible views as objectivist or subjectivist. Here, the objective character refers to an individual that prefers a tangible reality, usually having a scientific background and that is always focusing in results. On the other hand, a subjectivist individual is the one that focuses on individual perceptions, who usually comes from a social science background, and who focuses more on the means and not the ends. On the other hand, the variety of an evaluation party is made up of a series of elements such as the size of the group, the importance of the decisions taken upon the evaluation findings, the availability of resources, and the patterns of association. According to the strength of these characteristics the evaluation party can be seen as having high or low variety.

Taking into account these variables the authors developed a gridded type of framework having in the X-axis the approach to organizations and in the Y-axis the variety of the evaluation party, four possible types of approaches emerge as a result of combining this axes, these are: objectivist-high variety, objectivist-low variety, subjectivist-high variety and subjectivist-low variety. The framework is then used to discover four types of evaluation compatible with its four boxes, giving as a result the classification of the systems-resource based approach as an objectivist-high variety approach, the goal-based approach as an objectivist-low variety approach, the multi-actor based approach as a subjectivist-high variety approach and the cultural based approach as a subjectivist-low variety approach.

However, Gregory & Jackson along with other authors recognize limitations in their

framework. Gregory (1992, 1996) and Mingers (1992) criticize the structure of a framework that uses a gridded structure. Gregory (1992,1996) argues that categorizing methodologies in boxes unnecessarily restricts how the methodologies can be interpreted and hence the range of uses they can be put to, whereas Mingers (1992) considers that this type of categorization does not use enough variables to adequately reflect the complexity of the choices that people need to make in practice. Gregory & Jackson (1992) recognize two limitations of their framework. First, that none of the four forms of evaluation are explicitly oriented towards radical change, and second, that the exercise of power might influence the variety and approach to organizations of a group of people. Taking these limitations into account, Gregory (1997) explores appropriate methods of evaluation to use in coercive contexts. In order to do so, she first explores the forces of power present in organizations by reviewing the work of post-modernist authors such as Foucault (1976) on the issue. Given the definition of what power might be in the view of post-modernists, Gregory (1997) seeks to find a form of evaluation that may work well in coercive contexts. However, she argues that at first a form of evaluation for use in coercive contexts does not immediately suggest itself, and for this reason she follows the suggestion given by Flood & Jackson (1991) of using Critical Systems Heuristics (CSH) in simple coercive contexts as it emphasizes the empowerment of the stakeholders.

Taking into account the empowerment character of CSH, and the fact that it is more focused on what the organization ought to do than on what it is, Gregory recognizes some similarities in Stake's Responsive Evaluation and CSH including the focus on the identification of minority interests, the involvement of stakeholders, the making explicit boundary judgments and the empowerment of affected individuals (A deeper description of Critical Systems Heuristics and Responsive Evaluation is given in the previous chapters). For all these similarities, both approaches are regarded as complementary and represent a potential contribution to what can be seen as critical evaluation methodologies. Gregory proposes that Responsive Evaluation might be used to investigate what the bases of worth is within the organization and this information could then be used to fuel an exercise in CSH to determine what ought to be the basis for the determination of worth in the organizational sense, establishing an iterative mode of inquiry between the two. Although the introduction of an evaluation form to be used in coercive contexts represented a major advance from the initial work, this new development raised doubts from other authors, leading not only to a critique but to a reframing of this work. Boyd *et al* (2007) doubted that this type of methodology

could really work in the face of coercion mainly because although the approach surfaced minority viewpoints, it was still dependent of either debate between stakeholders or arbitration by a “higher authority”. Aside of this consideration, the authors also critique the fact that the development of the first framework and its further extension, used a grid form suggesting that “evaluation is made much more flexible, responsive and dynamic through the adoption of a systemic intervention approach than through the use of a contingency table” (Boyd, et al., 2007:7).

Gregory & Jackson (1992) conceptualize critical awareness in two ways, one of which is through the use of CSH in situations that have been defined as coercive. However, Boyd et al (2007) problematize this, by asking how a situation is defined as coercive in the first place. In order to give an answer to this question, the authors suggest that “systemic evaluation (as opposed to the use of a contingency table) must incorporate some explicit reflection, and preferably also dialogue, on what is to be evaluated and why, and whose views count” (Boyd, et al., 2007: 5). This is how a new approach to systemic evaluation emerges.

Systemic evaluation is based on creative design of methods and it emphasizes critique throughout the entire process with the aim of avoiding marginalization and seeking to make explicit those boundaries that are being considered in an evaluation and asking those involved in it whether these boundaries are justifiable or not (Boyd , et al., 2007). It also makes an explicit consideration of the internal and external relationships of an organization, exploring the consequences of these relationships as well. Systemic evaluation uses three of the forms of evaluation initially considered by Gregory & Jackson, namely; the stakeholder evaluation (multi-actor evaluation), the goal-based evaluation and the organizational evaluation (systems-resource evaluation), the culture form of evaluation is not considered because the authors thought it was not well developed. The three approaches may be used participatively in order to avoid marginalization.

Stakeholder or multi-actor evaluation, is based on the idea that organizations are primarily constituted by the interactions of social actors with their own unique subjective perspective. It assumes that the desires of the “external” and “internal” stakeholders are equally legitimate and establishes the role of the evaluator as being one directed towards assessing people’s levels of satisfaction as well as of minimizing the constraints over the organizations ability to satisfy its stakeholders’ objectives. Goal-based evaluation, views organizations as goal seeking, which need to be achieved and measured by means of quantitative measures of

inputs, processes, and outputs. Decision makers are viewed as rational beings capable of defining goals and the actions to achieve them, and being willing to adjust such goals and actions with the feedback received from evaluation. According to Gregory & Jackson (1992) management by objectives (Drury, 1985) is typical of goal-based evaluation: objectives are formulated and expressed in the form of performance indicators used to measure results (Boyd , et al., 2007). Finally, organizational evaluation or systems-resource evaluation, is undertaken with the survival of the organization in mind, which in extend requires adaptation or modifying external and internal environments. The evaluation focuses on the quality of the organizational processes that ensure it is in the healthiest “bargaining position” possible with the environment. In this form of evaluation, existing organizational processes are compared with a model of the ideal adaptive organization in order to generate ideas for improvement.

Figure 41. Relationship between the three approaches to evaluation.

Figure 10 shows how the three evaluation models are connected, it is expected that rather than choosing between the models, practitioners get to use all of them in a sequential form. However, the sequence is not strictly prescriptive, instead it should work as a general guide. The sequence to the ordering ensures a primary focus on the community of stakeholders whom the organization serves and interacts with, and then everything else flows from (Boyd , et al., 2007). By using each one of the approaches, the evaluators will be able to focus in different evaluative goals. For example, when conducting a stakeholder evaluation practitioners may focus on recurrent expressions of the same view, ideas for change or instances of outstanding good or bad practices that might signal the need for action. This can be achieved by exploring the *is* and the *ought* or by looking for recurrent expressions of the same views or ideas.

As can be seen the use of CSH and the development of Systemic Evaluation frameworks represents an advance from the practice of evaluation using the approaches in the first two

systems thinking waves. The use of the third wave approaches explicitly considers the implications of framing a situation in different ways and of establishing different boundaries for it. As Reynolds *et al.* (2012: 5) argue “a critical (third) systems thinking tradition of concern for boundary judgements invites the notion of designing purposeful engagement but also supports an explicitly ethical and political engagement with evaluation”. This also involves a critical analysis and reflection on the power issues that are often associated with boundaries, which determine what is relevant, legitimate or ignored.

In terms of the contribution that these approaches and particularly the use of boundary critique can represent for conducting evaluation, we see that a shift in paradigm needs to be made in that field. As Callaghan (2008:407) points out:

“The conventional evaluation, in making objectivist claims, purports to independence based on rigorous method but such evaluations frequently fail to be critical of the conceptualizations inherent in the policy process itself. This becomes an increasingly significant issue because of the institutional encouragement toward empiricism relying on an implicit acceptance of the dominant paradigm of traditional science. The very promise of social science begins with its ability to question the objectivist basis of such authority and, indeed, to reflexively recognize the impact of such constructions on the processes themselves (Callaghan, 2008)”.

Boundary critique represents an adequate approach to take evaluation in a new direction. Although they are not perfect in their application, the applications of the third wave of systems thinking in evaluation seek to address the power considerations that may surface in organizations as well as the possibility of marginalizing not only stakeholders but also the interests of each one of them. However, knowing when, how and why are we including different perspectives in an evaluation is of major importance as well. Multiple perspectives are generally included in the planning stages of an evaluation or by exploring the different perspectives of the stakeholders when analyzing results. However, we propose that using boundary critique throughout the different stages of an evaluation (planning, implementation, and analysis) may enrich the evaluative process even more.

A strength of the type of approaches explored in this section, is that they explicitly reflect, not only on the design of the evaluation, what and whom should be included, but also on what methods are more suitable for particular situations and why giving them a theoretical basis and a coherence throughout its application in the different types and stages of evaluation.

All of the authors approach evaluation by taking into account the different evaluation approaches that are present in the evaluation field, this represents an opportunity to not only

integrate systems thinking (particularly its third wave) and evaluation at a practical level but also at a theoretical level as well.

4.4.3 Frameworks for evaluation practice

Other authors have also carried out a relevant work for the practice of evaluation by using systems thinking. However, their work cannot be classified as belonging to a particular wave. For this reason, we will describe their developments in the following section.

The work of Gates (2016) aims at developing a framework to understanding the implications of using systems thinking and complexity science in conducting evaluation. It is important to take into account that although Gates recognizes a difference between what constitutes systems thinking and what constitutes complexity science, she analyzes the impact of the use of both fields in evaluation as a whole, suggesting that further research might be useful to conduct an analysis and gain a deeper understanding of the implications of each separately. The author identifies six core evaluation activities as part of her proposed framework, namely; supporting social problem solving, framing an intervention and its context, selecting and using methods, engaging in valuing, producing and justifying knowledge, and facilitating use. Later on, she explores how each one of these activities is affected when being explored from a systems thinking and complexity science perspective.

According to Gates (2016) the general contribution of performing these activities under a Systems Thinking and Complexity Science (STCS) approach can be summed up as follows. First, in order to fully take STCS into the evaluation practice, a shift from a linear predict-act-evaluate approach towards an iterative, adaptive, non-linear approach regarding social problem solving. Second, the consideration of multiple framings and the use of STCS concepts to choose one of these framings, gains relevance. A descriptive character towards the context surrounding an evaluation is no longer useful. Third, recognizing the partiality and the limitations of choosing and using particular methods becomes important, as this process needs to be done by giving consideration to the underlying assumptions of each one of them. Fourth, although Gates (2016) recognizes that the effect of STCS on judging the value of social intervention is not clear, she highlights the fact that some STCS approaches offer a process of comparing models which may offer alternative basis for making value judgments

Fifth, producing and justifying knowledge in the evaluation field has always been related to prioritizing knowledge of the effects, meaning that results and outcomes have always been put at the front when conducting an evaluation. However, STCS opens a door to open causal

relationships between different elements interacting in the system that is being evaluated. This means that the attention of knowledge production should not only be put on results but on the entire dynamic of the interacting elements including multiple sources of evidence and taking into account causal-relationships. Here, it is important to note that there are particular approaches that focus in a deeper way on comprehending the interrelationships of a system (e.g. Systems Dynamics), but these approaches may lack the tools to exploring other elements such as the inclusion of multiple perspectives, which means that the evaluator might need to be able to produce a mixed-method evaluation approach. Finally, Gates recognizes that STCS allows the opportunity to embed evaluation use into an intervention and/or agency by designing evaluation as feedback loops, social learning processes, or normative reflections and direction setting.

As can be seen, the work of Gates recognizes the importance of starting to conduct evaluation by using Systems Thinking and Complexity Science as these approaches open a new door to conducting evaluation in a completely different way from how it has traditionally being conducted. However, we will argue that some of the benefits Gates (2016) claims of working with STCS in the evaluation field cannot be generally attributed to all the approaches comprised in these two fields. First of all, as could be seen in the examples explored in the previous sections of this chapter, approaches belonging to the first wave of Systems Thinking might be very useful to understand the interactions of what is taking place in a system that is being evaluated. However, the same authors that have worked with this approaches recognize as major pitfalls the impossibility of working with different perspectives of stakeholders and also of having difficulties with reframing situations, as these approaches can be seen as an “instant picture” of a particular moment of the situation and cannot represent its evolution. We argue, that this is one of the key elements when conducting evaluation that has generally been left aside. Although we recognize the impossibility of including all the perspectives of the interested parties, it is important to at least include more than just the views of those who are commanding the evaluation. The inclusion of multiple stakeholders will not only bring more insight for planning, and focusing the evaluation but also will give the possibility of bringing changes to the current design of the program that will affect those whose voices are traditionally dismissed, and who are usually the directly affected by the program or policy that is being evaluated. On the other hand, being able to adapt to the real life and real time changes that might be taking place in an organization raises the possibility of what is being proposed to be useful.

Another framework developed to serve evaluation interests is the one proposed by Hummelbrunner & Reynolds (2013). In this framework, the authors relate the three systems concepts (interrelationships, perspectives, and boundaries) for the evaluation of a situation that were presented at the beginning of this chapter with three types of learning and three types of values that each one of them might be pursuing. In order to do so, the authors not only use the proposed systems concepts but also use a model of organizational learning based on the work of Bateson (1972), and Argyris & Schön (1978). Based on the model of organizational learning that includes a single, double and triple loop learning, the authors associate each of the three systems concepts with a specific loop of learning, expecting that the progression from single to double and triple loop learning lead to deeper and more sustainable learning.

In this sense, single loop learning will be focused on interrelationships primarily between the intervention and its effects, however the purpose of the intervention is not questioned. In the double loop learning, the assumptions underpinning an intervention can only be reflected if multiple perspectives are taken into account. In this case, the purpose and goals of an intervention can be questioned, as the situation can be framed in different ways. Finally, in the third loop learning, the focus is on the boundaries that are made with an intervention and its evaluation. This type of learning, does look at the power relations that determine the boundaries of an intervention and its evaluation, including the role of evaluators and evaluation commissioners.

At the same time, each of the learning loops is associated with a set of values. Single loop learning is related to instrumental values, this type of values inform evaluative measures regarding issues of efficacy and efficiency. Double loop learning is based on intrinsic value, they can include measures regarding issues of relevance and effectiveness. Finally, the triple loop learning is related with critical value, this is, value in reflecting on the rules and customs that govern dominant behavior and cognition patterns, they inform measures regarding issues of equity and emancipation. Figure 11 illustrates how the system concepts, the set of values and the learning loops are related through the framework.

Figure 5. A systems-based framework for rigor in evaluation.

The relevance of the framework not only lies on the relation of the three categories of concepts but also in the fact that it is useful to interrogate the coherence of the various components of an evaluation assignment. Hummelbrunner & Reynolds (2013) argue that reflections based on frameworks such as this one, can add rigor to evaluation practice. Addressing the appropriate value base can enhance the relevance, validity and credibility of evaluation.

4.5 Final Remarks

As can be seen, the use of systems thinking in the evaluation field represents a major shift from how evaluation has traditionally been conceived and conducted as it is commonly been related to conducting statistical analyses. The use of systems approaches allows evaluators to be more aware of how they are approaching situations to be evaluated and also give them a wider understanding of the forces that might be shaping a situation, forces that cannot be seen when using other types of approaches.

When using systems approaches in evaluation, there might not be wrong and right but how selecting the appropriate approaches can ensure getting the fullest picture of the situation and being able to conduct the richest evaluation as possible. We argue that in our view the best way of doing so is by using approaches of the third way of systems thinking, not only for in our opinion they are the most complete approaches in the three waves of systems thinking but also and most importantly, they have a critical character which is not only focused on understanding but also on questioning if what is being evaluated, how it is being done and the

methods that are being used are the most appropriate ones.

The use of this type of approaches lets us include different perspectives, and handle the differences among them as they are aware of the existence of power at organizational levels, and the need to address it in order to widen the evaluation impact as possible. The third wave approaches, are also in advantage compared to approaches in the other two waves in the sense that they are not only useful as diagnostic tools but they are useful to direct intervention and produce recommendations as well. One of the main strengths of the third wave approaches is their dynamism and ease when being used. This means that, they are adaptable not only to be used in a wide range of contexts and situations but they are also “friendly” with different types of users as no level of complex expertise is required to use them, making it easier for different stakeholders being part of an evaluation to apply and assess their results in a more direct manner. On the other hand, despite the different goals that might be sought through a program evaluation, its ultimate goal is social betterment (Marc et al., 2000). In this sense, social betterment constitutes an improvement which is why it is important to know in whose eyes the definition of improvement is being considered so that the impact of a program will not only be measured in relation to the interests of those who command an evaluation, but also considering as much as possible the interests of those who are directly involved with the conduction of the program. Ultimately Evaluation should then concentrate on understanding the capacities for interactions to promote improvement, as also evaluators establish both formal and informal communication, spaces for communication, and rewards for enhancing interaction (Reynolds *et al.*, 2012).

The fact that the relationship between systems thinking and evaluation has not been deeply explored (Gates, 2016), raises an opportunity for use to contribute to develop theoretically informed practices that will widen this relation and will give an opportunity for evaluators to use systems concepts in evaluation in a more comprehensive supported manner. Particularly, I would like to relate the systems concepts regarding boundary critique to the work that has been carried out with Fourth Generation Evaluation, by giving importance to the need of being reflective about how the evaluation is being framed throughout its different stages and how the presence of judgments cannot not only be attributed to judging the performance of a program but also to the different phases of its planning, implementation and feedback. The next chapter will describe how I pretend to relate both concepts, taking into account the theoretical review that has been shown so far in the preceding three chapters.

Chapter 5 – Contribution

The following chapter addresses how Systems Thinking ideas and specifically those of boundary critique can be beneficial for the evaluation field and particularly for the practice of Fourth Generation Evaluation. In the previous chapters I addressed how Systems Thinking and Evaluation fields have developed so far. For this reason, it will be assumed that the reader is familiarized with the general ideas and concepts that will be mentioned here.

Why Fourth Generation Evaluation?

Evaluation approaches generally follow a positivistic approach where the objective character

of the evaluation plays a central role, particularly shown in the external role given to the evaluator and the quantification of results. However, Fourth Generation evaluation represents a step forward in this matter as Guba & Lincoln (1989) sought to move beyond a positivist paradigm to a constructivist one which contrary to what is traditionally assumed in evaluation, possesses a subjective and inter-subjective character and gives a particular and important role to the evaluator throughout the stages of the evaluative process. Fourth Generation evaluation also takes a step forward in the inclusion of the stakeholders of the evaluand (what is being evaluated) in selecting the issues that need to be addressed by the evaluation, this is done primarily by means of the dialectic/hermeneutic circle. This form of evaluation represents a mix of two wider forms of stakeholder involvement approaches, namely participatory evaluation approaches (Guijt, 2014; Zukoski & Bosserman, 2018; Zukoski, Bosserman & Baechler, 2018; Chouinard, 2018) and collaborative approaches (Rodriguez-Campos, Rincones-Gómez & Roker; O'Sullivan). Other ways in which stakeholders could be included in program evaluation are represented by empowerment evaluation approaches (Fetterman, Kaftarian & Wandersman, 1996; Fetterman & Wandersman, 2018; Dugan, 2018; Fetterman & Ravitz, 2018). Although these three wider forms of evaluation have a particular theoretical base that distinguishes them from one another, they have as common trait the relevance given to the participation of stakeholders in at least one of the program evaluation stages. Although in general terms Fourth Generation evaluation constitutes a step forward compared to what is traditionally done in the evaluation field, I consider that some of its elements can be extended and studied under a different lens by approaching it by means of systems thinking concepts and particularly those related to boundary critique. A particular characteristic of this evaluation approach is that boundary critique is already present in it. However, this takes place implicitly by the way in which some of its methodological steps are conducted and by means of its guiding principles. One of the main tasks of this chapter is to make explicit the consideration of boundary critique in the stages where it is already present, as well as introducing this and other systems thinking concepts by means of reframing the methodology and some of its foundational conceptions.

For this reason, this chapter will focus on framing Fourth Generation evaluation from a systems thinking point of view, particularly oriented towards addressing boundaries and perspectives.

Why Boundary Critique?

Throughout the three systems thinking waves we have explored so far, the third wave is particularly important as it is directed towards addressing methodological pluralism, handling different perspectives, tackling power, and being aware of the importance of approaching a system of reference by means of an intervention and not only by observing it. This means that by intervening in the way an inquiry unfolds, an intervention might be able to modify a particular form of behavior so that its outcomes will be directed in a different direction from the one they were initially intended to.

By intervening in a system, one will always have a particular way of framing it due to the ideas, values, interests, and objectives that shape the way we see things. This is where boundary judgments appear, as a way of framing a situation, and by being reflective on the way such boundaries are set, boundary critique emerges. However, the importance of considering the way we frame situations does not only lie in the way we establish these boundaries but also in how by exposing them to the critique of others, they can be enriched, reshaped, moved, changed or dismissed; this is the essence of boundary critique. Although these concepts, boundary judgments and boundary critique, might have an abstract character, they are useful for different types of situations, such as program evaluation. I believe that these systems concepts can be useful for program evaluation not as a tool (as systems thinking ideas and developments are generally being used in this way) but as a way of conducting program evaluation and developing its theory, a way which can be very beneficial for the development of the field and for directing it in a direction which moves away from its positivist character and one in which qualitative information can have equal importance to the quantitative one. I believe that this qualitative character can serve as a basis to root the quantitative work that needs to be conducted.

Introducing systems thinking concepts at a theoretical level in the program evaluation field, by means of reframing Fourth Generation evaluation turns out to be beneficial given that the systems thinking concepts I seek to introduce as well as the methodology of this form of evaluation share similar foundational characteristics, making them compatible in theory and practice.

Given the previous considerations, I will address how the third wave of systems thinking, boundary critique (as well as the establishment of boundary judgments), can be used for reframing Fourth Generation evaluation and how its consideration can be useful to enrich the practice of the evaluation field beyond considering systems concepts as tools as has been

done so far, as well as contributing to the field in a theoretical way.

5.1 Reframing the Fourth Generation Methodology

The following section is directed towards addressing how different elements from systems thinking can be incorporated in Fourth Generation evaluation. In order to do so, I will start describing how systems thinking elements in the overall process of the Fourth Generation evaluation methodology can be useful and beneficial for its theory and practice. Afterwards, I will address how this evaluation approach implicitly makes use of boundary judgments, and boundary critique in specific stages of the methodology as well as how it does it. Finally, I will discuss the role of consensus and data collection in the approach and how they relate to these concepts as well.

5.1.1 The systemic character of Fourth Generation evaluation

In previous chapters I have addressed how systems concepts have been related to the practice of program evaluation. By exploring these connections, one is able to see that although both fields have had a similar development, most of the systems ideas that have been used in the evaluation field, have been directed towards using systems thinking methodologies as a useful way of conducting particular steps of an evaluative process and not for grounding their theoretical foundations. However, there are between three and four systems concepts that according to several authors' evaluators need to know (Imam, LaGoy & Williams (2006), Reynolds (2007), Hummelbrunner (2011), Reynolds *et al* (2012), Williams & van't Hof (2014), and Cabrera, Colosi & Lobdell, 2008). These are: perspectives, boundaries, interrelationships, and the system concept itself. The differences among these authors' views are explained in **Chapter 4** referring to Systems Thinking and Evaluation.

One can reframe Fourth Generation evaluation by identifying the presence of these concepts in different ways; multiple perspectives and boundaries represented by the conceptions of the stakeholders involved within the evaluation stages and the way in which they frame the situation, the interrelationships not only between the stakeholders but also between the issues, claims, and concerns that interest them, and the system as a whole by understanding and taking into account the contextual characteristics surrounding the evaluand and the evaluation, and which are at the same time affecting them and vice versa. These elements will be explained in more detail in what follows.

Given that Fourth Generation evaluation is based on responsive evaluation as well as

constructivism, as Guba & Lincoln (1989) express in their work and as it was explained in Chapter 3 of this document, one of its key components is the need to conduct program evaluation by taking into account the perspectives of those who have a stake in it. This means that it does not prioritize the way the client of the evaluation or the evaluation team seek to orient the evaluation. Instead, it places all the stakeholders in a similar position regarding the ability of designing the evaluation, based on the issues, claims, and concerns that all the involved consider as being relevant to be evaluated. An important element of these multiplicity of perspectives is that it not only eases the way in which the stakeholders can express their view points, by means of making constructions, but also by giving the same relevance to every construction. In that sense, prioritization of issues, concerns, and claims is not based on who makes them but on the direct implications of choosing one or another and on the results that can be achieved by doing so. By choosing particular viewpoints to be addressed, a boundary process takes place, as the evaluation starts to be bounded by the considerations of those involved in it. However, the key element that should be explicit is the fact that the consideration of multiple perspectives, does not only implies different ways of framing situations, but also on considering that the background from which these perspectives come have different elements that influence and shape them such as values, ideas, and beliefs. At this point is where perspectives and boundaries are particularly intertwined.

On the other hand, another important element is the role of the evaluator as a facilitator throughout different stages of the evaluative process that involve the participation of different individuals and stakeholder groups. When playing the role of a facilitator, Guba & Lincoln (1989) recognize the evaluator as having no special license, elite status, or superior power, nor is he warranted in exerting any special control, that is as being neutral. According to Rifkin, Millen & Cobb (1991), neutrality could be identified in two senses: first, impartiality, that is, ensuring against bias; and second, “equidistance”, which is to say ensuring that the parties participate “equally” within the process. However, taking into account such definition, the role of the evaluator could not be neutral given that even if the evaluator was in a different position within the evaluative process, it will always be biased in some sense, in other words it will always have a partial character. Given that partiality is understood as being grounded on biases, the stakeholders’ understanding, including the evaluator, will be partial as they will be biased by their own interests, goals, values, beliefs, perspectives, and distinctions they can make. Such biases will also lead to the formation of different

boundaries. Biases will be explored in further chapters of this research. On the other hand, the evaluator cannot be neutral in terms of ensuring an exactly equal participation of the parties involved in the evaluative process. Although one of the key elements of both, the original and the reframed form of Fourth Generation evaluation is the stakeholders' inclusion, this does not guarantee equality in their participation. Setting the adequate conditions for their participation does not guarantee it either, given that once again the issues which condition the human perception and behaviour come to the fore in the way in which they participate and understand situations as well.

Guba & Lincoln (1989), shift from a realist ontology to a relativist one. In their conception, there is no reality except that created by people as they attempt to make sense of their surrounds. At the same time, they also conceive truth as a matter of consensus among informed and sophisticated constructors, not of correspondence with an objective reality. In order to develop my proposal, I draw upon Guba & Lincoln at a methodological level given that my work has an epistemological character. This does not imply the adoption of a relativist reality, the rejection of the possibility of an ultimate reality, or even embracing solipsism. In order to take advantage of Guba & Lincoln ideas on Fourth Generation evaluation, one does not need to adopt a particular ontological position.

Guba & Lincoln (1989) recognize in their original development of Fourth Generation evaluation the importance of considering the stakeholders' interests, values, ideas, among other characteristics, as they are the ones shaping the issues, claims, and concerns of interest of those involved within the evaluative process. However, the way in which the authors propose to conduct the methodological development of Fourth Generation evaluation, does not explore how one individual's ideas, interests, values, and other characteristics may influence other individual's perceptions and how they may shift them. Instead, they focus on the individual contributions in the formation of constructions, given that the participation of multiple stakeholders does not necessarily imply an interaction between them. The proposed methodology, shifts from this conception towards one in which construction formation takes place simultaneously not only by involving different individuals but different stakeholder groups as well. Hence, the reframed proposed methodology implies a constant reflective process of setting boundaries around the different elements of its design and implementation, shifting between them and reshaping them until a particular boundary is reached. Such boundary will be the one that ultimately will shape the way in which the evaluation will take place in practice. It also becomes important not only to reflect over the implications of setting

a particular boundary but also over which are the assumptions, conceptions, and values that lead to the formation of particular boundaries. This is a process which at the moment does not take place explicitly in the Fourth Generation Evaluation approach.

An example of such elements are the “what” and the “who” that should be considered in conducting the evaluation. This reflective process implies the inclusion of new information which has the possibility of modifying the existing boundary, as well as putting the initial boundaries under the scope so that they can be placed in a position where they are critiqued. However, when using the term critique, this does not mean that they are targeted in a negative way so that their flaws might be put forward. Instead critique in a systemic sense implies that none of the constructions are taken for granted or assumed as an unalterable truth, so that they can be enriched and modified by putting them under a different scope from which they were originally conceived. On the other hand, according to Ulrich (1983) being critical means to become self-reflective in respect to the presuppositions flowing into one’s judgments, both in the search for true knowledge and rational action. This means that being critical in the reframed form of Fourth Generation evaluation is not only oriented outwards but inwards as well. This means that it is equally important and valuable to question not only other people assumptions but also one’s own assumptions. This would lead not only to the possibility of modifying other stakeholder boundary judgments but for reframing the own as well, as well as to trying to understand how and by what means they have been formed.

Ulrich (1988b) highlights the importance of considering boundary judgments in evaluation. According to the author, evaluation shows an inability to act effectively with regards to value judgments particularly because of their uncritical acceptance of the prevailing objectivism of science upon which most evaluation approaches rest upon. Ulrich (1988b) proposes the critical use of systems ideas in order to tackle this issue. When using system concepts in any real world issue or situation such as public policy and evaluation, a priori assumptions about what belongs to the systems and its environment, are made. These assumptions are defined by Ulrich (1988b) as boundary judgments. As I have stated earlier, boundary judgments as well as Fourth Generation evaluation, have a subjective character. This means that there is no right or wrong way of bounding but different perspectives in which this process can take place. A particular way of framing determines further ways in which problems and improvements are conceived, different boundaries bring to the fore different ways of appreciating improvement and defining problem situations. Boundaries are not only considered in framing the system but in considering the problem context as well and the consequences of bounding it in one

way or another. This is another element that is being introduced in reframing Fourth Generation Evaluation under the scope of system concepts and particularly those related to boundary judgments and boundary critique. Guba & Lincoln propose four different purposes with which an evaluation could be carried out. However, two of those purposes do not take into account the contextual elements surrounding the evaluand and the evaluation. As such, those evaluation purposes should not be considered when conducting a reframed Fourth Generation methodology as this would dismiss the systemic character of the proposed methodology given that they do not take into account the evaluand and the evaluation context for which they do not explore the relationship between wider systems [contextual framing] and the sub systems [evaluand and evaluation]. The interaction between the wider systems and the sub-system will be determined by the way in which the former is framed; given that it could be framed in multiple ways, one can have multiple wider systems to explore.

Churchman (1979) calls for an unfolding process of the different interpretations and their life-practical implications for those who may be affected by a decision. The unfolding process can take place in a "monological" (self-reflective) or in a "dialogical" (discursive) setting; in both cases the value of the process depends on the extent to which the true concerns of all the stakeholders, especially of those not involved but possibly affected by the decision in question, are considered by those involved (Ulrich, 1988b). The process of unfolding furnishes *the critical counterpart to the sweep-in process* and its endless quest for comprehensiveness. The critical intent of the unfolding process becomes evident here: while there is obviously no end to the sweeping in of still other groups of citizens who ideally would benefit from a systems design, the idea of the unfolding process, as Ulrich understands it, is not to avoid a boundary judgment regarding the client question but rather to develop awareness and consensus about the "true" purpose of the project, i.e., its criterion of success. Hence the ideal client is not "everybody," as the sweep-in idea might lead one to expect; such an answer would beg the question. In the unfolding process, the client question aims precisely at the fact that not everybody can be the planner's client and that no conceivable decision will serve everybody's needs (Ulrich 1983, p. 393). This constitutes an example of how boundary judgments are helpful when addressing stakeholders and their interests. When reframing Fourth Generation evaluation by these concepts one can identify the necessity of explicitly considering boundary judgments in the way in which the "who" and the "what" of the evaluation are being addressed and framed. This is also relevant in understanding that framing an evaluative process implies that total comprehensiveness cannot totally be

achieved, given that not all the identified stakeholders may be relevant for the evaluative process, for which since the beginning the issues, claims, and concerns of their interest will be left outside and also because all the information about the evaluand may not be relevant for achieving the purposes of an evaluative process, for which although it may be kept for future incorporation [just like the marginalized stakeholders], it will not be taken into account. Total comprehensiveness also turns out to be unachievable given that an infinite expansion of the boundaries of the evaluation is not attainable or desirable.

On the other hand, when referring to the systemic character of the underlying rationality of this form of evaluation, I refer to its ontological, and epistemological elements. The ontological assumption of this type of evaluation differs substantially to those developments from the three previous generations of evaluation. The original form of Fourth Generation is characterized by having a subjective relativistic character which is based on the impossibility of achieving an objective truth about reality. This assumes that reality is “an act of construal and is independent of any foundational reality” (Guba & Lincoln, 2001). Denying the existence of an objective truth, also means rejecting the so called subject/object dualism, meaning that the observer has a direct influence on the way it is observing and perceiving an object. In this conception lies one of the common elements between systems thinking, particularly its second and third waves, and this form of evaluation. According to Midgley (2000), this way of conceiving an inquiry [by rejecting the subject/object dualism] will strengthen the critique of so-called value free science so that the values flowing into observation can be made more visible. The consideration of these values, as well as other elements that shape our perspectives, are the basic considerations of the Fourth Generation epistemology.

The epistemological element of the Fourth Generation evaluation is directed towards a “transactional subjectivism”, in Guba & Lincoln (2001) terms. This means that assertions about “reality” and “truth” depend solely on the meaning sets and degree of sophistication available to the individuals and audiences engaged in forming those assertions. Assertions about reality and truth have a bounded character as the way in which an individual conceives reality differs from the way somebody else does. The main reason for this to happen is the fact that the way in which one shapes reality, this is how we frame it, is constrained by the way in which our perspective is formed by the influence of different elements such as our values, beliefs, ideals, personal goals, desires and others. These elements work as a source of information for us to make a particular judgment regarding an issue. As a result it turns out to

be impossible to have a unique way of appreciating and understanding things, meaning that there is not a right way of doing so either.

On the other hand, interrelationships can be introduced by two means, one directed towards the stakeholders and other directed towards the contextual elements surrounding the evaluation. The former is related to the need of understand how the stakeholders play a particular role in the evaluand [the program to be evaluated] as well as the relational dynamics among the different stakeholders. These relationships are a key element in understanding how particular constructions are formed, as well as understanding the dynamics that take place in the evaluand and that could ultimately determine and explain how issues, claims and concerns unfold and how the evaluand operates. On the other hand, when referring to contextual interrelationships, it is important to recognize the interrelationships between the contextual characteristics and the dynamics of the evaluand, as well as how such contextual characteristics may condition the way in which the evaluation is designed and put in practice. It is important to recognize macro and micro contextual characteristics; the former are important for understanding the conditions in which the evaluand operates but they may not have a direct impact on it, and the latter are relevant for the evaluand as they directly influence it.

Reframing Fourth Generation evaluation by using these concepts is very useful for the evaluation theory and practice given that it shows how an evaluation could become systemic in its methodological structure and not only in its practice.

5.1.2 Boundary judgments, boundary critique and the reframed form of Fourth Generation Evaluation

As I have described, one of the most important elements of the Fourth Generation rationale is the denial of the subject/object dualism. This assumption poses four considerations for the practice of this type of evaluation that will be explored in what follows.

First, this means that each stakeholder, namely client of the evaluation, evaluation team, and those having a stake in the program, has a different perspective regarding the issues related to the program under study and its evaluation. This means that each individual as well as each group of stakeholders will frame the inquiry in different ways. Each of these ways of framing is influenced by the values, ideas, objectives, and interests of the individual. The general implication of this statement is that there is not a unique way of perceiving the program or the

task that will be pursued by means of the evaluation; instead it highlights the multiplicity of boundary judgments attributed to the different stakeholders, as well as the necessity of choosing a way to prioritize or select some of them.

Second, although an appreciation of the program might be considered as an observation, an objective way of observing the inquiry is impossible. In that sense, observation transforms into an intervention, that is, one which not only “observes” but which also intervenes having a direct effect over what is being observed as well as how it is being observed.

Third, being aware of how decision making processes takes place, as well as of the role of those intervening in the evaluation, will have an impact on the way the evaluation is designed, implemented, and oriented towards producing recommendations that will be more likely to lead to generating changes. In that sense, by framing the situation in a way in which those involved in it as well as their considerations are taken into account, might derive in producing recommendations that are more likely to be put in practice than if the evaluation is done only by considering only what is relevant for the client and the evaluation team. For this reason, it is so important to involve multiple perspectives by means of involving different stakeholder groups throughout the evaluation stages. However, this does not mean that every single issue, claim, and concern proposed by the stakeholders will be taken into account, given the restrictions surrounding the theory and practice of an evaluative process, and the fact that some of such issues, claims, and concerns may not be relevant for the interests with which an evaluative process is being conducted, even though the participation of the different stakeholders is expected to take place after such prioritization takes place.

The fourth and last consideration is that by including stakeholders in the different stages of the evaluation, an attempt is made to give the same relevance to the perspectives of all those included, that is all the stakeholders seen as relevant for conducting the evaluation although this might not be reflected in the prioritization of issues to be evaluated as not all the issues considered by the stakeholders could be tackled at once.

5.2 Consensus in Fourth Generation evaluation

Just like Guba & Lincoln (1989), Ulrich (1988a, 1988b) considered that if a rationality is dialogical, plans for improvement should be considered as normatively acceptable to all those participating in a given dialogue (Midgley, Munlo & Brown, 1998). In this case, improvement lies not only in the set of recommendations that can be given by means of an evaluation or the need to reduce the gap between a desired state and the current one, but also

in the different ways those involved in it influence its design. Something that is normatively acceptable refers to something that is being generally acceptable as a standard which can be translated in the achievement of consensus regarding a topic (in this case, about the selection what and who should be considered to conduct an evaluation). However, Midgley (1992, 2000) argues that the consideration of different stakeholders and issues related to them in a systemic intervention implied the existence of different boundary judgments grounded in different sets of values. According to Midgley (2000) and Ulrich (1983), systemic intervention refers to intervention that pursuits the ideal of comprehensiveness. However as achieving absolute comprehensiveness is impossible, an adequate methodology for systemic intervention should facilitate considerations of issues of inclusion, exclusion and marginalization by promoting reflection on boundaries, as well allowing theoretical and methodological pluralism. An evaluation could be considered as a systemic intervention, not only for seeking comprehensiveness through which improvements could be recommended, but also because by including boundary judgments and boundary critique reflection over boundaries can be achieved, as well as achieving theoretical pluralism by means of considering an arrange of theories that can enrich the design and practice of Fourth Generation evaluation, and methodological pluralism by using more than one methodological approach not only to enhance the practice of evaluation but also to enrich the comprehension of the pertaining issues of the evaluation. In the particular case of Fourth Generation evaluation, methodological pluralism can be achieved by addressing the situation by means of quantitative and qualitative methods, as well as including systems thinking methodologies to be used in specific stages to achieve defined purposes.

The differences that emerge by the consideration of different stakeholders and their constructions, are the ones that might lead to the appearance of both conflict and consensus within the evaluation rather than outside it. This conflict is the one that shapes the selection of a particular boundary for framing the element of the evaluation that is being addressed. This is another element that although is not exactly aligned to what Guba & Lincoln (1989) state about consensus, is the base of the negotiation and the dialectic processes that take place within the evaluation phases. Moreover, this element prevents the appearance of conflicts later on that might produce undesirable consequences, escalation processes, and decisions not reached by consensus.

Whether Guba & Lincoln are aware or not about this, according to Fisher, Ury & Patton (1991) a negotiation process can be defined as a ““back-and-forth communication designed to

reach an agreement when you and the other side have some interests that are shared and others that are opposed (1991) ” Several elements can be identified in this definition. First, is the need to have two or more sides so that a negotiation between them can take place, in the particular case of Fourth Generation evaluation, this refers to the stakeholders who have some stake in the program, whether they are clients of the evaluation, part of evaluation team, or are being affected (positively or negatively) by the program. Second, is communication, in order to be able to negotiate, the sides need to develop some form of communication, whether it is written, verbal or by other means and an effort should be made on the way the message not only leaves the sender, but also on how it reaches the receiver and how he/she interprets it. Third, is the fact that those involved in the negotiation have interests, some of which might be similar and others do not. In Fourth Generation evaluation, this is represented by the issues, concerns and claims relevant for the different stakeholders. As each stakeholder has a different role on the program and in its evaluation, as well as a particular background (values, interests, ideas, and goals) that shapes their perspectives, they will have differences in these claims, issues, and concerns as well as on the way they prioritize them. These differences are the ones may be solved by means of the negotiation. This is how consensus can effectively be achieved. However, in order to do so, dialogue and/or negotiation needs to be used when necessary.

One can also identify these elements in a dialogical process. However, the main difference is that in the latter the speaker can stand in its initial position without having the necessity of moving from it. A pure dialogue, is conducted only with the aim of achieving mutual understanding and not for generating some sort of agreement. The latter is what is needed in evaluation. In that sense, dialogue as a process in which decisions and not only understanding are needed, could be more precisely approached in the way Franco defines it; “jointly creating meaning and shared understanding between participants [of the dialogical process] (2006, p.814)”. In the dialogical hermeneutic process that takes place in the different stages of the Fourth Generation evaluation, one can identify an implicit shift from what is traditionally understood as a dialogical process to a negotiation or a dialogical process in terms of Franco (2006). According to Guba & Lincoln (1989) the dialogical hermeneutic circles that take place in this form of evaluation are based on the constructions made by different stakeholders involved in them. However, as a new group of stakeholders is added to the circle, the constructions that were previously formed by other stakeholders are put under the scope of the new ones by means of a critique. By doing so the new stakeholders not only

critique the constructions of others but also bring their own to the fore. In that sense, the initial stakeholders can reshape their conceptions, and the new constructions brought by the new stakeholders can be critiqued as well, turning the interaction into a cyclical one. A process that only ends when some sort of consensus or agreement is achieved. However, this interaction and reshaping of constructions somehow can be seen as a negotiation in which a particular side decides or not to give on its initial conceptions. This process acquires a negotiation character not only for reshaping constructions, but also because of the need to achieve consensus on particular issues which are primarily represented by “what” and “who” will be included in the design and implementation of the evaluation. In that way one can see that the achievement of some type of consensus first moves from the differences between the stakeholders, differences that might generate conflict, and that need to be managed by means of a negotiation.

5.3 Relevance

As can be seen systems thinking elements, particularly those related to boundary critique are very useful for framing and conducting a Fourth Generation type of evaluation. The main benefit of using boundaries in an explicit manner is that at plain sight the stages of the evaluation seem to be easily conducted and consensus presents itself as easily achievable. However, there are several processes that need to be carried out and taken into account in order to reach this point. It is of particular importance realizing that although one might count with few or several stakeholders, the main difficulty in conducting participatory evaluations that give a more active role to the stakeholders is that it is difficult not only to make them aware of what matters to them, but also on how what matters to one or another can be equally useful to carry out an evaluation. Another important element, is that by using boundary critique one as evaluator can be aware of the decisions made throughout the evaluation design and implementation, as they not only shape the way in which evaluation was conducted but also how results could be interpreted and recommendations could be produced. In that sense, since the conception, the entire evaluation process is interrelated and affected by the decisions made throughout the evaluation.

Regarding the establishment of boundary judgments, boundary critique and marginalization, one can recognize a different way in which this take place, as by having several boundaries, although not in a primary or secondary character, this can also take place.

The next chapter will explain in detail the reframed form of Fourth Generation methodology

and each one of its steps.

Chapter 6 – Contribution Methodological Design

The evaluation I put in practice, takes into consideration systems thinking and evaluation approaches in a single methodological intervention in a way in which a particular evaluation approach is reframed by explicitly introducing systems concepts in it. The entire methodological design and the reframing of the approach will be explained in what follows.

Regarding evaluation theory, I decided to base my work on the use of Fourth Generation evaluation, an evaluation approach consisting of 12 steps. Although this form of evaluation

sets the ground for the evaluation conducted in practice, in order to introduce the systems concepts at design and implementation levels, several changes needed to be made. First, the 12 steps were not conducted in the same way as stated by Guba & Lincoln (1989). Instead some of the steps were “merged”, others were modified, and some were not explicitly conducted. Second, systems concepts are traditionally included in an inquiry in an implicit manner which makes it more difficult not only for the evaluator but for the reader as well to be aware of them. In this case, aside of using systems concepts in the methodological design as a whole and on specific Fourth Generation evaluation steps, they also represented the inclusion of two additional steps from the original ones. More details over the modifications done to the Fourth Generation methodology will be given further in this chapter. Third, although Fourth Generation evaluation sets the ground for conducting the evaluation of the Hermes Program, several of its steps were not conducted in the same way as they were proposed. Instead I designed a specific way of using them which will be illustrated in what follows.

The 12 original steps to conduct Fourth Generation evaluation are²:

1. Initiating a contract with the client or sponsor commissioning the evaluation.
2. Organizing to conduct the evaluation.
3. Identifying the stakeholders.
4. Developing within-stakeholder-group joint (collaborative, shared) constructions via the hermeneutic/dialectic circle process.
5. Testing and enlarging within-group constructions: this is achieved by introducing new or additional information and by enabling group members to achieve higher levels of sophistication in dealing with such information.
6. Sorting out resolved claims, concerns and issues: these claims, concerns, and issues are those in which consensus has been achieved and can be of historical interest as well as being considered in other settings.
7. Prioritizing unresolved claims, concerns, and issues: in some cases the evaluation might not be able to address all the unresolved claims, concern, and issues as they might be too many or because the evaluation resources are too narrow to work with all of them, for this reason prioritization of issues needs to be achieved.
8. Collecting information referring to the unresolved claims, concerns, and issues.
9. Preparing the agenda for negotiation.

² These steps are explained in chapter 1 of this document.

10. Carrying out a negotiation.
11. Reporting.
12. Recycling.

However, the reframed methodology consists of the following steps:

1. Establishing a relationship with the client.
2. Understand the context in which the program takes place.
 - 2.1 General context.
 - 2.2 Specific context.
3. Identifying adequate stakeholders (the “who” component).
 - 3.1 Phase 1: Evaluator-client.
 - 3.2 Phase 2: the rest of the stakeholders.
4. Reflecting upon the stakeholders’ selection.
5. Developing, testing, and enlarging constructions.
6. Identifying and prioritizing relevant issues, claims, and concerns to be addressed.
7. Reflecting upon the boundaries about “what” should be addressed by the evaluation.
8. Identification and design of adequate tools for data collection.
9. Data collection.
10. Analysis of results.
11. Discussion and sharing of findings to generate recommendation with the stakeholders.
 - 11.1 Technical component.
 - 11.2 Explanatory component.
12. Generating recommendations.
13. Producing reports.
14. Recycling*

As can be seen, this methodological design considers two additional steps regarding the reflection over the selection of boundaries and not conducting the final step of the original evaluation, that is Fourth Generation evaluation. However, the final step of Fourth Generation evaluation, consisting of recycling several issues, claims, and concerns is not considered in this evaluation mainly because as the evaluation was conducted within a limited time frame it cannot be extended in the same research. However, this does not mean that further research could not be conducted by considering the issues, claims, concerns and results from this research.

The activities conducted in each of these steps will be addressed in what follows.

6.1 Establishing a relationship with the client

According to Guba & Lincoln (1989) this phase should include the identification of the client, the identification of the evaluand (what is going to be evaluated) which is based on the different interpretations and constructions of those involved and the establishment of purpose for evaluation. However, in the reframed methodology, this step is directed towards establishing the client-evaluator relationship, defining the purpose for the evaluation, and setting an initial step in defining the evaluand and in recognizing possible relevant stakeholders for the conducting the evaluation. Focusing on a different arrange of elements than the one initially proposed, would make it easier to gain clarity with regards to how the subsequent stages of the evaluation should be addressed.

Guba & Lincoln (1989), recognize two elements to be included in establishing the purpose with which Fourth Generation evaluation could be conducted, these are: having a summative or a formative character, and establishing if it is directed towards assessing merit or worth. However, a major difference in the reframed form of Fourth Generation evaluation, is that in order to attribute it a systemic character, it necessarily needs to be aware and take into account the contextual elements surrounding the evaluand and the evaluation. For this reason, and taking into account the definition of merit and worth, at this stage the reframed Fourth Generation methodology will be focused on assessing worth. However, it can still be conducted with formative or summative purposes.

On the other hand, the definition of the evaluand in the reframed methodology also differs from the original conception proposed by Guba & Lincoln. In order to introduce this difference, it is important to recognize two types of evaluand: a general and a narrow one. The general evaluand corresponds to the program itself. Both methodological developments, the one proposed by Guba & Lincoln and mine, have as their main goal the evaluation of a program. However, this does not mean that the evaluation will be directed towards evaluating the program as a whole. For this reason, a narrow evaluand emerges. The narrow evaluand corresponds to the actual section, area, issue or subject of interest for the evaluative process. In the proposed reframed evaluation methodology, the identification of the specific evaluand does not take place in a single methodological step. Instead, it is a process that takes place throughout several stages of the evaluation, given that it does not only involves the evaluator, but the participation of the client and other stakeholder groups as well. For this reason, this

evaluative step could be considered as a way of setting an initial boundary over the issues related to the selection of the specific evaluand of the evaluation.

Setting an initial boundary over “who” is considered to be a relevant stakeholder for the evaluation is something that could take place at this stage or further stages of the evaluation. This depends primarily on the way in which the client and the evaluator establish a relationship. Given that the evaluation methodology has a participatory character, although the boundary over “who” should be involved in the evaluative process could change in the future, involving those stakeholders that have already been recognized as relevant in the evaluation is useful for identifying the contextual factors affecting the evaluand and the evaluation.

On the other hand, when referring to establishing a relationship with the client, the evaluator might be in a position, where a formal written contract might not be possible. However, there are other figures which could play an intermediary role to facilitate the communication and activities that need to be conducted between the two parts. This intermediary role is key in the success of the evaluation and in gaining credibility of the process and the results.

6.2 Understanding the context in which the program and the evaluation take place

In the original methodological design, after establishing the relationship with the client and defining the purpose of the evaluation, comes the logistical organization of the tools and processes needed to conduct the evaluation as well as that of the external factors affecting the program and the evaluative process. Two modifications regarding the original methodological design are proposed in the reframed methodology. First, the establishment of the logistic needs of the evaluation, needs to be conducted throughout the evaluative process, as at this point of the evaluation the only thing that is clear is the purpose of the evaluation and not what is going to be the focus of it. Second, two types of external or contextual factors could be identified as relevant for the evaluation, these are: the specific and the general context.

The general context corresponds to the political, economic, social, technological, environmental, and legal factors surrounding the program and its evaluation, also known as PESTEL variables. These factors are important to understand the general conditions in which the program was created and in which it operates. For this reason, they should be taken into account when designing and implementing the evaluation. On the other hand, the specific context corresponds to the particular variables surrounding the program, such as competing

programs (these are the ones operating in the same geographical area or programs directed towards tackling the same type of issues as the evaluand), pressure groups such as gangs and drug dealers, local government and their policies, the local community in which the program is being developed and so on. Unlike the general context, these factors may have a direct visible impact on the operation and development of the program on a regular basis, whereas the general context is more likely to condition the way in which the program is designed. Furthermore, the importance of these factors not only lies on comprehending how they affect the program and the evaluation but on being aware and developing ways of addressing them in the evaluation design. On the other hand, understanding and taking into account the specific and general contexts in which the program and its evaluation takes place is part of conducting a systemic type of evaluation in which the main aim is to build a picture by which it would be possible to understand the relationships between the general and the narrow evaluands and what surrounds them, as well as reflecting over the implications of such relationships.

Both of these stages, establishing the contract with the client and organizing to conduct the evaluation, are key elements for enriching the perspectives about the program and how the evaluation could be properly designed. This is because establishing the contextual characteristics relevant for the evaluation is not a task conducted exclusively by the evaluator but one in which other stakeholders could be included as well. Even more, these steps define the first boundary by which the inquiry is started to be framed. This framing will be reshaped and customized through the following stages of the evaluation.

6.3 Identifying adequate stakeholders (the “who” component)

In the original Fourth Generation evaluation methodology, the stakeholder’s identification has two components; their identification and their inclusion in the evaluation. The former is usually conducted by the evaluator whereas the latter is conducted by means of a negotiation including the identified stakeholders in which they must justify their participation in the evaluation. At this point, the evaluator acquires the role of a facilitator for carrying out the negotiation among the other stakeholders. I consider that the process could be conducted in a different way given that in my opinion stakeholders should not be demanded to present their case in order to be considered as being relevant for the evaluative process. In my reframed methodology proposal, the identification and selection of stakeholders takes place in a different way in two separate stages.

The first stage involves a joint work between the client and the evaluator on identifying who they consider to be a relevant evaluator of the program. In doing so, they set an initial boundary over “who” they consider as being a relevant stakeholder for the evaluative process (This could take place in Stage 1 of the evaluative process). It is important to note that although a stakeholder may play a key role in the operation of the evaluand, it may not be equally important or necessary for the evaluative process of the specific evaluand. For this reason, when talking about the relevance of the stakeholders I am referring to their participation in the evaluative process. Once the client and the evaluator have identified a set of relevant stakeholders, is the turn of the identified stakeholders to draw their own boundaries over the “who” component of the evaluand. The second stage then consists of developing their own boundaries, and discussing the role that each one of the stakeholders plays in the program. The inclusion of multiple stakeholders in this decision making process is very important because they have different perspectives over the same issue. For this reason, their inclusion turns out to be enriching compared to the situation in which the task is conducted only by part of the stakeholders. Their inclusion is also relevant because each one of the stakeholders as well as the evaluator addresses other stakeholders and their role within the evaluand differently. Once the boundaries have been set, it is the task of the evaluator to choose or frame a particular boundary. This will be explained in more detail in what follows.

Boundary critique turns out to be very helpful in decision making processes related to the “who” and “what” components of an evaluative process. Each one of the available evaluation approaches follows different ways of pursuing these tasks. Some delegate this decision over the evaluation team, others follow the explicit instructions based solely on the interest of the client (which is the one who commands the evaluation) and others do it by following some sort of participative approach. In the original design of Fourth Generation evaluation, “who” should be considered for participating in the evaluative process is decided by means of a negotiation between different audiences, whereas the selection of “what” should be considered as pertinent and relevant to be evaluated is done by means of the formation of constructions by the relevant stakeholders. The way in which both processes are reframed by the consideration of boundary critique, as well as how they both unfold, will be tackled in what follows.

In the original design of Fourth Generation evaluation, the decision over “who” should be included in the evaluative process is grounded on the recognition of stakeholder groups as agents, beneficiaries and victims of the evaluand. It is the evaluator’s task to make such

recognition as well as that of uncovering additional stakeholder groups that may be relevant for the evaluative process which were not initially considered. In order for a stakeholder group to be considered as relevant for the evaluative process, it must justify why its participation as well as the issues, claims, and concerns of its concern would be useful for the evaluative process. According to Guba & Lincoln (1989), the main reason for selecting stakeholders by means of a negotiation is that it constitutes an attempt to place equal levels of power around those involved in it, by giving the possibility of participating for all the identified stakeholders. Nonetheless, providing conditions for equal access to participation does not necessarily means that every stakeholder will for sure participate in it. This has also to do with the individual skills, knowledge, personality and other traits associated individually to each of the participants which also have a direct impact in the negotiation³.

Although the way in which Guba & Lincoln propose to choose “who” should be part of the evaluative process may seem innovative in terms of achieving some sort of participation as well as that of a negotiation, I consider that this process could be improved not only by taking into account the stakeholders’ opinions but also by being aware of the implications of including certain stakeholders and not others. I propose to reshape this task as follows. First, the decision making processes related to “who” should be included in the evaluation will now be taken in two separate stages involving different stakeholders. Given that each individuals as well as every stakeholder group will bring different types of knowledge, interests and viewpoints the consideration of different stakeholders is valuable. For this reason, the decisions over “who” should be included in the evaluative process should not rely solely on the evaluator.

As the relationship between the client and the evaluator is established in the first stage of the evaluative process, it is their task to recognize those stakeholders who may be relevant for it. The first stage of establishing “who” should be part of the evaluative process is then carried out between the client and the evaluator. Their interaction, may lead to the establishment of the first boundary judgment around the “who” component of the evaluation. If the decision making over the “who” component was effectively taken jointly between the client and the evaluator, it is very likely that boundary critique would have taken place as well given that each one of them would have departed from different boundaries. As shown in Figure 12, the

³ This issues have been tackled by different authors such as Habermas (1970) by means of the ideal speech situation and Ulrich (1983) by introducing the Critical Systems Heuristics. Although efforts have been made to tackle these issues, they are always present in the relations between stakeholders, even more in conducting negotiation processes.

client and the evaluator set their own boundary or boundaries over the stakeholders to be considered in the evaluative process. By means of a dialogical process, an overlapping region denoted by * emerges. This region corresponds to the stakeholder groups both of them recognized as being relevant for conducting the evaluative process. Once this is achieved, the process moves on to the second stage regarding the “who” component of the evaluation.

Figure 6. Boundary critique in stakeholders' selection.

Taking the overlapping region as the output of the interaction between the client and the evaluator, the second stage would be directed towards involving the recognized stakeholders in the evaluative process. These stakeholders will carry out a framing process as well, a process which may lead to widening or narrowing the resulting boundary of the interaction between the client and the evaluator. The difference between the role played by the stakeholders here and that proposed by Guba & Lincoln, is that they are not put in a position in which they need to justify themselves or others but one in which they are asked to address the situation as being active participants of it and for which they are capable of building a “picture” of the dynamics that take place within the evaluand and not the evaluation. It is worth noting that stakeholders are not asked about particular issues of concern for the evaluative process. Instead, they are asked about their perceptions of the evaluand. Given that this second stage involves the interaction of the different stakeholders, it will lead not only to the establishment of a new boundary over the “who” component, but also to the fact that boundary critique is necessarily present again. The idea of this process is that by involving representatives of the different relevant stakeholder groups’ one may be able to form a second boundary.

Once there are two available boundaries, the one formed between the client and the evaluator

and the one of the stakeholders, it becomes the evaluator's task to decide which one of the boundaries will be the most appropriate one to be taken into account in the evaluative process. It may be the case that both boundaries are the same for which there would not be an additional decision to be made. The reader may be wondering why this decision should only be taken by the evaluator or evaluation team. The reason for this is that the necessary knowledge for the evaluation's general methodological design resides in the evaluator and not in the other stakeholders, for which the evaluator is the only one that could explore the implications of framing the evaluative process in one way or another, and also by addressing the constraints that may have an impact over this decision and which other stakeholders are not aware of.

6.4 Reflecting upon the stakeholders' selection

After comparing the identified groups of stakeholders, as well as understanding the role and the interrelationships with which those stakeholders are part of the program, the evaluator can proceed to reflect if it is adequate or not to change (expand, contract, or reshape) the initial boundary of the considered stakeholders. Although the reader might think that this process should not be considered as an independent step from the previous one, I would argue the contrary. The main reason for considering this as an independent and relevant step is that by consciously doing this, not only the evaluator but the evaluative process can benefit by explicitly addressing the consequences of a decision over the design of the evaluation. The reflection over the boundaries is the responsibility of the evaluator or the evaluation team for three reasons. First, each stakeholder will bring different forms of knowledge to each one of the stages in which they get involved in the evaluative process. As the other stakeholders of the evaluand have already been involved in setting different boundaries, the knowledge they possess regarding the evaluand is no useful in this stage, as the evaluator is the only one who is aware of issues and constraints that may affect the operation of the evaluative process. Second, the reflection is over the performance of the evaluation and not over the direct operation of the program. This means that it is necessary to address the implications of framing the evaluative process in terms of the relevance of the participation of stakeholders with regards to the purpose with which the evaluation will be conducted. Third, if other stakeholder groups were to participate in the reflective process and ultimately in the decision making, conflict is very likely to arise.

6.5 Developing, testing and enlarging constructions

The decision making processes over the “what” component of the evaluative process, are linked to the way in which the “who” component has been framed. Both processes are linked as the “what” component is grounded on the issues, claims, and concerns of the considered stakeholders of the evaluand. Having a particular set of stakeholders as being part of the evaluation design represents a way of bounding the situation, not only by the issues, claims and concerns they bring to the fore but also because having left out other stakeholders means that there are issues, claims, and concerns that will never be tackled as they may not be represented in any way by the interests of the stakeholders who are being considered.

Constructions are the base upon which Fourth Generation evaluation grounds its methodology and as such they are built several times and in different ways throughout it. In Guba & Lincoln (1989) original conception of the methodology, constructions are made in two separate stages. At first a single group of stakeholders makes its own constructions regarding the issues, claims, and concerns relevant to them, and after they are tested and enlarged by placing them into contact with other groups of stakeholders who have the task of forming new constructions as well as critiquing the ones formed by other stakeholders. However, in the reframed methodology, the process takes place by following such an intertwined dynamic that it should not be seen in a separate way. Although at first the process might be carried out within a single group of stakeholders on a wider scope, the process unfolds in a very similar manner as by generating, testing and enlarging these constructions reflection takes place. When forming constructions within a single group of stakeholders, one would believe that it is very likely to have a low variety in terms of the constructions made by those involved. However, every stakeholder and actor is influenced by its own particular set of values, ideas, interests, and beliefs. This implies the formation of particular constructions (this does not mean that several constructions may have a similar character) as well as that of other individuals who may be participating in the dialogical process. For this reason, the formation of boundary judgments as well as that of boundary critique takes places throughout the whole process. The role of the evaluator is directed towards setting the conditions and being a facilitator for and in the dialogical process.

In order to move from a within the same group of stakeholders level to a more general one, consensus over a single or a set of constructions needs to be achieved. In order to do so, at first every stakeholder of the group is entitled to establishing its own boundaries, which refers to framing the issues, claims, and concerns regarding the situation in a particular way by

means of the formation of a particular construction. Thereafter, the actors within the stakeholder group proceed to share and critique the constructions made by others. However, when talking about critiquing, this does not refer to pinpointing flaws and deficiencies of others' constructions. Instead, this is directed towards enriching the process of boundary formation which is directed towards what is considered as being adequate, under the perspectives of those considered to be relevant to the evaluand and to be included in the evaluation. In that way constructions could be tested, by being discussed with others, and reshaped by considering the elements and the position of other members of the stakeholder group as well. Hence, a process in which boundaries are reshaped, constrained or expanded takes place. This process now replicates in groups considering not only considering representatives of a single stakeholder group but other stakeholders in the same group as well where the constructions upon which consensus was achieved within a group of stakeholders will be under scope. The reshaping process that may take place when exploring different constructions is another important element which was not recognized by Guba & Lincoln in their original methodology. Given that constructions are being formed, critiqued and tested between different individuals (within or between stakeholder groups), the dialogical character of the process makes it possible of an individual's interests and viewpoints to be directly influenced by other individuals ideas, opinions, values, and interests over the same issue.

The way in which the general process unfolds at the upper level is almost the same as the one conducted at the lower one in terms that the same process of generating constructions, testing them and reshaping them are necessary. However, the main difference between the two processes is that in the former, part of the elements that limit the perspective of the participants belong to the same type of actor, which is a characteristic that somehow bound the way in which they see things. When involving other groups of stakeholders, the process will be enriched in a wider scale compared to the one originally conducted as this stakeholders aside of bringing different issues, claims, and concerns, bound the original ones in a very different way from that of who propose it at first. This means that variety is a key element for constructions formation.

After conducting both processes, within stakeholder groups as well as between stakeholders, two outcomes are generated. First, are those issues, claims, and concerns which are no longer of interest for the evaluation as there is a consensus among the participants which means that they no longer have a problematic character that needs to be explored. Second, are the issues, claims, and concerns in the form of constructions upon which the stakeholders did not

achieve any consensus that still have a problematic character and should be explored. However, in some cases these issues, claims, and concerns may be very extensive to be tackled by means of a single evaluation. For that reason, the next step needs to be conducted. It is important to clarify that in this stage consensus refers to achieving some sort of agreement in the way in which a particular issue, claim, or concern is seen. This means that these issues, claims, and concerns do not have a problematic character. Instead, those issues, claims, and concerns upon which consensus is not achieved are the ones that might be considered to be addressed in the following steps of the methodology. That is, by means of posing them as central points for conducting the evaluation.

As each of the stakeholders has a different way of viewing the situation, this implies that their viewpoints need to be critiqued and put under scope by the other members of the audience so that disagreements might emerge and by this negotiation process they can achieve some type of agreement over *what* should be included and what should be left out. This is an iterative process which will lead to the establishment of a new boundary and the possibility of the emergence of a marginalization process. This process takes place from the initial stages of the evaluation. However what varies is “who” is involved in it and “how” the boundary concerning the “what” issue is reshaped throughout time.

Marginalization will only appear if both boundaries are different, meaning that there will be a marginal area between them. In that case, if the selected boundary is the primary one, this is the narrower boundary, those considered to be relevant by the secondary boundary will be left out of the evaluative process. This process could be interpreted in two different ways; one related to marginalization itself and one related to exclusion. The former refers to diminishing the relevance given not only to the participation and consideration of the stakeholders throughout the evaluation (that is the *who* component), but also to those issues that are relevant to them, in that sense *what* they see as relevant is given a lesser value that *what* is considered by other stakeholders. The latter, refers to completely excluding the stakeholders from the development of the evaluation not only by not giving relevance to their issues of interest, but also to explicitly excluding them as individuals. Although a particular boundary might be chosen in this stage, after developing other stages of the evaluation design and execution, the evaluation team might be aware of a necessity of changing the boundary which was initially selected. This is one of the benefits of using boundary critique in the evaluation design, as those involved in it might be conscious of the implications of framing the situation in a particular way or another. Although maybe not immediately, they will be aware of which

are the desired changes in the evaluation and how by reframing it they can be achieved.

Given that not all the issues, claims and concerns exposed by the stakeholders can be included, the final process of the dialectic process represent a marginalization over the set of issues that were initially in the reference system. Although the final selection is narrower than the initial reference system, in many cases it still covers too many elements to be tackled by means of a single evaluation. For this reason, the next step in the evaluation process is that of prioritizing the issues, claims and concerns that will be directly tackled by the evaluation as well as giving an appropriate use to those elements which got to be resolved by means of the dialectic circle. The latter are dismissed meaning that they will not be taken into account for conducting further stages of the evaluation as they are no longer relevant to it.

6.6 Identifying and prioritizing relevant issues, claims, and concerns to be addressed

Boundary critique is not limited to what should be addressed by the evaluation. The next step of the Fourth Generation evaluation consists of prioritizing which issues, claims and concerns should be tackled by the evaluation. As Guba & Lincoln (1989) express it, in many cases a single evaluation cannot address all the relevant issues, claims, and concerns raised by the stakeholders. For this reason, prioritization needs to be done, so that the evaluation can focus in what is considered most important. There is not a single way of prioritizing these issues, claims, concerns. This means that although it can be carried out by means of a dialectic hermeneutic circle, it does not necessarily needs to be done in this way.

Whether the prioritization is conducted by means of a dialectic circle or not, it will have a boundary critique as well as a marginalization process implicit in it primarily because in this case prioritization does not mean that those issues, claims, and concerns that are being marginalized are being completely discarded; rather it refers to the impossibility of tackling all of them in the same intervention and having the opportunity of keeping them on hold in order to use them if necessary in the future so that the current intervention could be enriched by them or not. In that case, these issues, claims, and concerns acquire a shifting character in terms of the value given to them as they might be seen as less valuable or irrelevant at present but acquire a more prominent character in further stages or in future interventions.

In the previous step of the evaluation, the stakeholders formed a series of constructs that were critiqued and reshaped until a consensus over a set of constructions or a single one of them was achieved. The set of constructions over which consensus was reached, represents the initial boundary from which possible prioritization needs to be done. Prioritization does not

necessarily take place if the resulting set of issues, claims and concerns is narrow enough to be tackled in a single evaluation.

However, when prioritization effectively takes place, not all of the issues, claims, and concerns upon which consensus was achieved will be considered in the final scope of the evaluation. In terms of the establishment of boundaries and critiquing them, these initial set of issues, claims, and concerns represent an initial single boundary which will later transform into a secondary boundary as prioritization implies drawing a narrower boundary, namely a primary boundary (Figure 13).

Figure 73. Boundaries of the issues, claims, and concerns to be tackled by the evaluation.

The primary boundary corresponds to those issues, claims, and concerns that will be directly addressed by the evaluation, whereas the secondary boundary corresponds to the initial claims, concerns, and issues from which the ones belonging to the primary boundary were selected. In this case, it is evident that a marginalization process occurs as some of these issues, concerns, and claims will necessarily be left outside of the evaluation and as the primary boundary will always be selected (at least when having too many issues, claims, and concerns). This type of marginalization is not associated to the exclusion of a particular group of stakeholders. Instead, it is focused solely in the issues, claims, and concerns that may be tackled by means of the evaluation as at this stage the considered stakeholders are not diminished or excluded in any way from the intervention. It might be the case that the stakeholders from which the issues, claims, and concerns are marginalized, might feel themselves as being marginalized as well as it will look like their interests are not being taken into account. But it is necessary to take into account that the marginalization of these issues, claims, and concerns, is done in a joint manner which means that all those involved in it have the possibility of “defending” the inclusion of their personal considerations. In some cases, some type of accommodation between the considered issues, claims, and concerns, might be achieved and in that sense, marginalization does not necessarily takes place.

There are two possible approaches when considering and establishing boundaries. The first one, refers to the case by which boundaries are clearly set without the need of reflecting over

several boundaries at once. A second case taking place in interventions, is the one in which possible boundaries are explored before setting them, however this is less common than the first approach. In this case, the primary boundary does not exist at the same time as the secondary one, instead it is the result of the process which explicitly consists of discarding elements considered by the secondary boundary that will be addressed by the evaluation. In that sense, the process is not based on the selection of a predetermined boundary but on drawing a new one which necessarily needs to be narrower than the former one. Marginalization occurs as a result of drawing a narrower boundary as this means that a set of issues, claims, and concerns will not be part in any way of the current evaluation design and implementation.

However, in this case marginalization has a special connotation. Although some issues, claims, and concerns might be left outside the current evaluation design and implementation, this does not mean that they will be completely discarded or ignored, instead they can still be used in the future in two ways before or after completing the current evaluation. If this is done before the evaluation is ended it might be done with the purpose of gaining insight about the issues considered in the initial evaluation or to link its results with elements that were not initially considered, a process that might not only be expanding the chosen boundary but also its systemic character by considering interrelationships that were not initially taken into account. On the other hand, when considering these issues, claims, and concerns after completing the first evaluation, this might be done with the purpose of giving a different scope to the evaluation or with the aim of broadening what was initially done. Whether these issues, claims, and concerns are used later on or not, its relevance lies in their availability to be used in a future time frame.

Although selecting a particular set of issues, claims, and concerns, might be related only with the selection of a primary boundary, such as the selection of “who” and “what” (in broader terms) will be included in the evaluation, puts pressure on the different ways several groups of stakeholders frame situations, the way they need to be tackled, as well as how (they consider) prioritization needs to be done. The fact that all these steps of the Fourth Generation evaluation are conducted in a participatory way, pose a challenge in the appearance of conflict. Guba & Lincoln (1989) tend to see the negotiation and the dialectic hermeneutic circle as processes where consensus is achieved and can be used as a signal to move forward to the next step of the evaluation. In this sense, I decided to approach Fourth Generation Evaluation by considering the definition of consensus given by Checkland (1990) and not the

one given by Guba & Lincoln who were the actual developers of this form of evaluation. According to Checkland (1990), consensus is given by the accommodation of the different interests presented by those involved in the different stages of the evaluation.

6.6.1 How to decide what to include and what to marginalize?

Decisions over what to include and what to marginalize are based on six considerations, three of which are an original proposition of mine, and the other three are based on Guba & Lincoln's work on establishing criteria for prioritization. The first one is given by the first step of the evaluation design. In such step, the evaluator and the client defined the main purpose with which the evaluation was going to be conducted. This purpose works as a guiding light for directing the evaluation design, and work as an initial boundary for what should be included in it. However, by including the issues, claims, and concerns of other stakeholders the initial purpose seeks to be expanded in a way in which the evaluation could be enriched and its central purpose could be supported and widened. The way in which such issues, claims, and concerns of the stakeholders are taken into account are related to the way in which they could support such central purpose. Two criteria are considered at this point: if the issue, claim, or concern has a direct connection with the evaluation purpose defined between the evaluator and the client, and if the issue, claim, or concern is linked to another issue, claim, or concern surrounding the purpose defined between the evaluator and the client.

Figure 8. How to choose the “what” component of the evaluative process..

Figure 14 shows the possible boundaries relevant for choosing issues, claims, and concerns to

be taken into account in the evaluation design in relation to the purpose established by the client and the evaluator. On the other hand, Figure 15, shows how the boundary related to the initial purpose defined between the evaluator and the client must be widened, whereas the boundary related to the proposed issues, claims, and concerns of the other stakeholders must be narrowed.

Figure 9. Prioritization of issues, claims, and concerns relevant for the stakeholders.

It is important to clarify that although there will be a set of marginalized issues, claims, and concerns, this does not necessarily mean that they will be completely discarded. Instead, this means that they are not relevant at the moment or for conducting a particular evaluation or stage of it. Given that the evaluation will produce a series of results and findings, the discarded issues, claims, and concerns may be useful to explore the understanding about the behavior of such results and findings in the future. Given that different issues, claims, and concerns are formed by different stakeholders, they not only represent different perspectives over the same topic or situation but also different standing points over which the perspectives are grounded. These differentiations are very useful for building a richer picture in conducting the evaluation as they may lead to multiple complementary findings.

Second, is the fact that some issues, claims, and concerns may seem more severe, problematic or of more concern than others, making them to be a priority to be tackled as soon as possible by the evaluation. Other issues, claims, and concerns may be furthered tackled by means of other evaluations or in further stages of the same evaluation.

Third, there is another important issue related to how such boundaries are set, such issue is related to power issues and to the motivations and preferences of the evaluator. Although the evaluator strives to have a neutral position with regards to the decision making processes about how to guide the evaluation, the truth is that its motivations, preferences, and values will be relevant in how this decision making process is conducted given that he or she can choose to favor particular issues, claims, and concerns based on its personal preferences. Although I attempt to standardize this decision making process by grounding it on the linkage

between the issues, claims, and concerns, and the central purpose of the evaluation, this does not necessarily eliminate the possibility of basing the decision making process on personal interests. An additional measure by which this bias could be minimize is by means of conducting such process involving more than one individual or by following an extended verification process involving other evaluators or other stakeholders aside of the evaluator and the client. By these means, although it is not certain, it is less likely to ground decision making on the personal interests of the evaluator.

Another way by which this could be minimized is by ascribing power to knowledge itself and on the possible outcomes that may be obtained as a result of exploring certain issues, claims, and concerns and not others. However, it is likely that the evaluator values the most issues, claims, and concerns he/she considers will be more useful for certain topics than others particularly his own predictions over the issue. This is particularly related to forms of cognitive biases of the evaluator. Power and biases in evaluation are addressed in more depth in Chapters 10 and 11 of this document.

One of the critiques of the original methodological design of Fourth Generation methodology, was the role of the evaluator. In such conception, Guba & Lincoln (1989) argued that the evaluator must attempt to become an advocate in order to redress imbalances in knowledge between stakeholders. However, conceiving power in terms of relational dynamics, reinforced forms of knowledge are a result of such dynamics, for which there are not real power imbalances but particular power relations that let to certain forms of knowledge. By this means, the evaluator does not need to become an advocate for any stakeholder group but instead needs to understand how knowledge, in the form of issues, claims, and concerns was generated.

The three prioritization criteria grounded on Guba & Lincoln's (1989) work are related to: the possibility of resolution of an issue, claim, or concern, the possibility of achieving an action compromise although dissensus might continue, and the centrality to the value system of one or more of the stakeholders. This is, given that issues that rest upon major value positions are the most important ones to deal with. These additional criteria have a practical orientation given that they are directed towards addressing issues in terms of the future capacity of being responsive to them.

6.7 Reflecting upon the boundaries about “what” should be addressed by the evaluation

Although most of the previous steps include reflection as a key element, being able to make a

final reflection about everything that has been chosen so far is important. This is the last stage of the evaluation where reflection can be conducted with the purpose of modifying what has been selected to be included in it. After this point, although reflection keeps being a fundamental element it can no longer be directed towards the evaluation design. Instead it could be directed towards future improvements for the evaluand or the evaluation.

The reflection taking place in this step needs to be directed towards addressing the possible consequences of conducting the evaluation by having as central aspects the elements chosen in the previous steps, if the evaluator or evaluation team have the adequate resources to focus on these elements, which are the constraints limiting the evaluation and the data collection regarding these topics, as well as which tools would be appropriate for data collection. As such this type of reflection should ease the decision of whether moving forward with what has been chosen or instead reframe the situation in a way in which the evaluator would be able to address particular issues with the available resources (human, monetary, time, etc.).

This step is particularly important to the evaluator, as it will be the most affected, whether it is in a positive or negative manner by the way in which the evaluation is directed as this has as a direct implication over the need to adapt to the particular context that each situation poses.

6.8 Identification and design of adequate tools for data collection

Once the evaluator is clear about its limitations as well as of what he/she wants to tackle, the next step is designing the tools to do so. Design is not necessarily a step that needs to be conducted in the sense that the desired tools might already have been designed by someone else. However it is necessary at least to identify which are the external available resources that could be helpful for the evaluator and its team to achieve its goals. To be able to design or choose appropriate tools gives the evaluator an opportunity of customizing tools for its specific needs. A helpful concept for making such decision and design, is methodological pluralism. This concept corresponds to the use of adequate methods according to need in conducting evaluation by using methodological partitioning as the technique for interrelating such methods. This is, the sectioning and recombination of parts of distinct methodologies complex research issues (Midgley , 1990). In order to do such thing, the evaluator must be aware not only of the different methodologies available but to the needs he/she wants to fulfill by means of the data collection, the different contexts surrounding the situation or situations to be tackled (considering each context one at a time), and the constraints

surrounding the data collection. In that way, he/she will be able to select or design more effective tools.

There are several considerations regarding methodological pluralism that need to be taken into account. First, according to Midgley (1990), pluralists, in this case evaluators, need to recognize and address the incompatibilities between the methods and perspectives they want to use in order to avoid theoretical contradiction between them. This is achieved by harmonizing the assumptions of traditional isolationist theories. Flood (1989), described the need for pluralism to be methodological incommensurable and theoretically commensurable. This means that they need to be adequate for different perceived situations while being separately defined at a methodological level, as well as seen as complementary at a theoretical one. Second, the use of multiple methods in a single intervention implies interrelating them systemically so that critical reflection is allowed. This is important particularly because the different chosen methods need to be relevant for all the defined contexts that might be present in a single situation. Midgley (1990) claims that the main advantage of methodological partitioning is that, when two or more approaches are to be used in an issue with several related contexts, the resultant methodology will prove a more useful source of information than the use of a single established method in isolation. The work of Midgley (1990) and Midgley & Floyd (1990) are examples of methodological pluralism in practice.

6.9 Data collection

After choosing “who” and “what” will be part of the evaluation and prioritizing these issues, the next phase of the Fourth Generation evaluation is collecting information regarding the chosen issues, claims, and concerns to be explored. This type of evaluation is different from most of the other evaluation approaches because it gives more relevance to qualitative results than to the quantitative ones. This is done because the purpose with which information is collected is for it to be added to the existing constructions, in order to lead to reconstructions, as well as having appropriate ways to appreciate and use the new information. However, the use of quantitative data is equally valid to the use of qualitative one. In that sense, Fourth Generation evaluation uses as its primary source of evaluation the design of new tools as well as previous information that has been collected and the previous experiences of the evaluation team. Part of what is done is to explore the constructions formed in dialogical processes to see if it is possible to use them to improve one’s understanding.

Although this might seem to be very helpful for the development of this type of evaluation, I consider that the most appropriate way for conducting this type of evaluation is by mixing qualitative and quantitative methods, where the evaluation can move back and forth between the two types of methods so that the evaluation practice can be enhanced. This process might take place in two different ways: by pluralism or by triangulation. The former is more prominent in Critical Systems Thinking whereas the latter has been widely accepted in the evaluation field. Pluralism refers to the use of diverse methods, whether they are qualitative or quantitative, in achieving particular goals. Triangulation, refers to cross referencing the information from different methods in order to enriching and making the results of the evaluation more robust. In my personal opinion, both ways of mixing methods could be achieved in a single intervention as particular methods could be used for achieving particular goals, which results are later on used in shaping further stages of the evaluation and selecting other methods in pursuing a further exploration of previous results in a different manner. In that sense, both results are somehow connected and are used to enrich the process as a whole and not only for the particular goals for which they were developed and conducted. On the other hand, this means that each particular method serves a specific purpose in one or several stages of the evaluation. An example of mixing methods can be found in Midgley's (2000) evaluation of Diversion from Custody for Mentally Disordered Offenders.

So far the evaluation has focused on its qualitative component as it has been working with the criteria of those participating in the evaluation design to determine the orientation the evaluation should have. However, at this stage those issues which could not be addressed by means of these negotiation and dialogical processes, need to be tackled in a different way. These issues, claims and concerns do not need to be solved in the sense that they are not seen as a problem. Instead the differences upon which they remain unresolved make reference to the differences of perspectives regarding a particular statement. Given the subjective character of the situation, one cannot expect to achieve an objective answer for the issue, instead what is sought is gaining comprehension about the general perception over a particular issue. For this reason, I consider that the qualitative character the evaluation has had so far, needs to be supported by means of quantitative methods.

However, in order to quantitative methods to be really useful in practice and helpful for stakeholders, Midgley (2017) considers that at least one of two things (or both) need to happen. First, the evaluator needs to secure agreement on what should be measured so all concerns are addressed, and second the evaluator needs to include the issues, claims, and

concerns, of all perspectives, or even go beyond them. This second issue is key as when using quantitative methods, it is better to widen the boundaries rather than narrowing them. By widening these boundaries, one is most likely to conduct a more enriched and useful evaluation that gives us a more complete picture of the situation. When addressing a wider arrange of issues, claims, and concerns, one is more likely to minimize conflict than to enlarge it.

Being able to use quantitative methods in the collection of information, gives the possibility not only to the evaluation team but also to all the stakeholders involved in the evaluation to gain a comprehension of the general behavior of the program. A program, as I will illustrate in further chapters of this document, might be so broad that in order to achieve representativeness, one would need the participation of a large proportion of those involved in it.

6.10 Analysis of results

This is as step which was not considered in the original methodological design proposed by Guba & Lincoln (1989). However, I propose it as there is technical knowledge that the evaluator may possess that other stakeholders of the evaluand do not have in order to be able to translate data into useful information for the evaluative process. By these means, for which it becomes necessary to determine the usefulness of the participation of other stakeholder groups. Once the evaluator has conducted all the necessary tests to collect data, it is time to interpret the gathered information. The analysis will vary according to the type of information that was gathered in the previous step. Producing results differs from discussion of findings in that although the stakeholders' participation is encouraged throughout the evaluative process, this step is very likely to be conducted by the evaluation team. As I have described in previous steps, different stakeholders (including the evaluator and the client) have different types of knowledge that are useful for conducting some evaluation stages and not others. For this reason, given that the analysis of results has a technical component, the evaluator or evaluation team are the ones likely to conduct this task.

Although other stakeholder groups may not be involved in producing results, this does not means that the findings will be imposed over them. Instead, the following stages are focused on exploring such findings and in producing recommendations between all the relevant stakeholders of the evaluative process.

6.11 Discussing findings and generating recommendations

Even though discussing findings and generating recommendations are two separate steps in the methodological design, in practice they can take place jointly for this reason I will first address them theoretically and then I will present how I consider the activities related to them can be conducted in practice. Both of these steps are similar to what Guba & Lincoln (1989) propose as preparing for the negotiation and carrying out the negotiation.

6.11.1 Discussion and sharing of results to generate recommendation with the stakeholders

In traditional forms of evaluation this task is assigned to the evaluator as he/she is the one who has the adequate knowledge for interpreting results. Although at first glance this might be true given that the evaluator might have an extensive statistical knowledge useful in the case of quantitative evaluation, or an ability to recognize verbal patterns for the case of qualitative evaluation, at the end the real interpretation not only depends on that. At this stage, two phases might be recognized; a technical and an explanatory one. The technical component is related to the analysis of the results by giving meaning to what was obtained in the previous stage. For example, if one has a percentage, the technical component will give the interpretation of what that percentage means and the implication of its value. On the other hand, the explanatory component is directed towards giving meaning in the particular context of the situation to the technical component. This component is the most important of all as it is the one that locates the results in the particular context surrounding the evaluand and the evaluation. In order to achieve better results in the explanatory component, the participation of the same stakeholders that participated in previous stages of evaluation is required so that they can participate in giving meaning to these results. The technical and the explanatory components are elements introduced in the proposed reframed methodology, given that none were considered in the original design. Involving such components reinforces the systemic character of the evaluation given that in order to explore results and proposing recommendations one needs to locate both within the real life situations in which they are taking place. This may lead to a more appropriate understanding of what is really happening.

On the other hand, Guba & Lincoln propose to conduct this stage by means of the hermeneutic/dialectic circle. However, I propose to conduct it in a different way which only keeps the dialogical component. It follow a similar structure to the one proposed to be followed in construction formation. This will be explained in what follows.

6.11.2 Generating recommendations

Generating recommendations is not a step in the original design of Fourth Generation evaluation. However, doing so might be important to produce better results for the future. The previous step was directed towards giving meaning to results by including the participation of the stakeholders of the evaluand. However, getting to solve particular issues, claims, and concerns should not only limit to that. Instead, there are always recommendations that could be adopted so that the evaluand might have a better performance in the future, or in the case where resolution was not achieved it might be achieved in the future by adopting some changes.

6.11.3 Discussing findings and generating recommendations in practice

As I have stated previously, in practice it is very likely that these two methodological steps will be conducted jointly. A way by which they may be achieved, follows the structure shown in Figure 16.

Figure 10. Steps for sharing findings and producing recommendations.

The description, proposition, and construction structure allows the evaluator and the other stakeholders (including the client) to have a back and forth relation regarding disclosing findings and generating recommendations about them. These steps, are surrounded by the contextual restrictions, as well as the evaluation constraints that restrict the proposition of “adequate” recommendations. In this process, the evaluator will be allowed to present the evaluation findings (description), while the other stakeholders get the chance to propose improvement ideas or changes (proposition), and to consolidate, reshape or make new ones jointly (construction). This design would tackle one of the critique of the original Fourth Generation design related to knowledge imbalances, given that all the stakeholders (at least by means of their representatives) are having access to the same knowledge with regards to the perform of the evaluand, the evaluator does not have the need of serving as an advocate of particular stakeholder groups, and the client is not having the authority of imposing decisions over other stakeholder groups.

Reshaping, consolidating and forming new recommendations is itself a process in which

boundaries are present as well. Although it may be less evident that in issues related to the “who” and the “what” components of the reframed Fourth Generation design, primary and secondary boundaries are present given that recommendations are being produced in two separate stages and it is not possible to keep all of them as they are proposed. By considering the appearance of such boundaries, marginalization emerges as well given that the possibility of addressing some recommendations later without the need to completely dismiss them is still possible.

6.12 Producing reports

After conducting the entire evaluation process comes the generation of reports which describe the processes that have taken place in the evaluative process. Particular attention is directed towards describing the constructions made by means of the hermeneutic/dialect circle in each of the stages in which it was required.

6.13 Recycling

Guba & Lincoln (1989) call recycling to the use of the unresolved issues, claims, and concerns that might be brought about the appearance of new information, so that they might be readdressed in order to try to solve them. However, recycling should not only limit to the unresolved but also to the discarded. In the initial stages of the evaluation, several issues, claims, concerns and maybe stakeholders might be discarded by the way in which the evaluand and the evaluation is bounded. Both of these elements, the “who” and the “what” which were originally left out, could be retaken into further stages of the program, the evaluation or even in a new evaluation not only for them to having some sort of relevance but also because in that way the comprehension regarding the evaluand will increase.

6.14 Final considerations

As can be seen, the Fourth Generation methodology has been reshaped not only by introducing systems thinking concepts but also by proposing different ways of conducting steps from the way in which they were originally proposed. With respect to the original design, two additional steps have been added, whereas others have been kept but conducted in different ways. Figure 17 shows a summary of the methodological design steps followed in the reframing of the Fourth Generation evaluation under the consideration of boundary critique and systems concepts.

Figure 11. Methodological design of Fourth Generation evaluation by considering boundary critique and systems concepts.

Besides of showing the stages of the methodology, there are several considerations that need to be addressed regarding Figure 17. First, although there are two specific steps directed towards reflecting upon the “who” and “what” of the evaluation, throughout the stages 3 to 11 boundary judgments and boundary critique are implicit in the activities that are carried out. Though this does not necessarily imply that the evaluator is aware that conducting such reflections it is a process that would be very useful when being done consciously. Second, the logistic arrangements related to organizing and coordinating the tools, materials and stakeholders necessary to conduct the different stages of the evaluation do not only take place in its initial stages as it was originally proposed by Guba & Lincoln (1989); instead it is conducted throughout the evaluation. This is because there are several stages in which the participation of stakeholders, as well as the conduction of negotiation and the hermeneutic/dialectic circle takes place. Third, the original methodological design of Fourth Generation evaluation consisted of 12 steps; reframing the methodology implied the inclusion of additional steps, two of which are related to boundary judgments and boundary critique, and another one related to the generation of recommendations. On the other hand, certain steps of the original methodology were grouped as I consider that they should take place in a

joint manner. Fourth, as I am only reframing the methodology, its ontological and epistemological foundations are kept as they originally were.

Figure 12. The methodological structure of the reframed Fourth Generation evaluation approach.

Figure 18 shows how the steps of the proposed methodology relate to each other, several considerations are to be explained. First, the methodology described above has been presented in a linear way for organizing and structuring purposes. However, this does not mean that in practice the methodology necessarily needs to follow the same order. Second, given that the methodology has a systemic character, several steps may be taken in order to conduct a single further step. Third, the contextual characteristics surrounding the evaluand are relevant for conducting all the methodological steps that take place once the context has been explored. Given that the study of the context is identified in step 2 of the methodology, conducting step 1 does not necessarily depends on such consideration. Fourth, given that Figure 18 is oriented towards presenting how the methodological steps relate to each other, several activities and concepts such as the logistical arrangements and boundary judgments which are relevant throughout the evaluative process are not taken into account. Fifth, those who are using the methodology in practice can go back and forth between each one of the proposed steps as much as they need to.

The next chapter will explain how each of the steps was implemented in the evaluation of the Hermes Program.

Chapter 7 – The Hermes Program

Up to this point I have addressed the theoretical foundation upon which my work is grounded. However, from now on I will describe how these theoretical concepts were put into practice by evaluating a program developed by the Bogota Chamber of Commerce called Hermes Program, oriented towards teaching alternative resolution techniques to children in vulnerable areas in Bogota, Colombia. In this chapter I seek to give an overview of the Hermes Program, how it was developed and how it operates so that the reader may form an idea about the evaluand.

7.1 Bogota Chamber of Commerce

The Bogota Chamber of Commerce (BCC) dates back to the year 1878, when a group of visionaries comprised of José Manuel Restrepo, Ramón del Corral, José Camacho Roldan and Jorge Holguín, set out to foster the creation of companies, defend the interests of the city's business owners and entrepreneurs, and to become an advisory body and government consultant (Camara de Comercio de Bogota, 2018a). The BCC defines its higher purpose as follows:

“We are allies in turning your entrepreneurial dream into a reality, helping make it sustainable, and generating shared value, in order for all of us to strengthen the Bogotá-Region, with more and better companies (2018)”.

This definition goes in hand with the three principles by which it is governed, these are: prosperity, shared value and governance. Prosperity corresponds to a state of growing and inclusive well-being achieved through the abundant and well-balanced creation of public and private goods and services, for all of the community. Shared value refers to enterprise practices which increase business competitiveness, while improving social, environmental and economic conditions in the communities they operate in. And finally governance corresponds to the ability by different stakeholders to work in a collaborative fashion, in order to manage their own development.

The BCC is divided into five dependencies focused on different issues related to the operation of the BCC. These dependencies correspond to: public registries, legal, business strengthening, competitiveness and shared value, and the center for arbitration and conciliation. The latter is the one in which the Hermes Program was created. The following section, will explain with some detail what this dependency is.

7.1.1 Center for Arbitration and Conciliation

The Center for Arbitration and Conciliation (CAC) is a branch of the BCC established in 1983 as a response to the need of Bogotá's entrepreneurs in resolving conflicts in relation with their activities (Cámara de Comercio de Bogota, 2017c). However, throughout the years, the scope of the CAC has extended beyond that of entrepreneurs to tackle the needs of common citizens in terms of conflict resolution. In order to do so, the CAC has opened different centers of operations as well as tackling conflict by Alternative Dispute Resolution methods (ADRs) such as arbitration, legal conciliation, insolvency, mediation, qualified experts, amicable settlement, equity conciliation, enterprise coexistence, community coexistence, and student coexistence according to the needs of the client. The Hermes

Program and the programs that preceded it were developed to attend the needs of audiences related to school environments that presented a need of addressing conflict.

The next section will address important issues regarding the Hermes Program, its creation and its methodology.

7.2 Alternative Dispute Resolution

Alternative Dispute Resolution (ADR) is a broad spectrum of structured processes, including mediation, arbitration, negotiation, and conciliation, which does not include litigation though it may be linked to or integrated with litigation, which involves the assistance of a neutral third party, who empowers parties to resolve their own disputes (Law Reform Commission , 2010) .The idea is that ADR is voluntary and either party is able to withdraw at any time. The proposed solution does not have to be accepted. These concepts will be briefly discussed in what follows, placing particular attention to conciliation as this method is the central topic of the Hermes Program.

Among the different types of arbitration, one can find ad-hoc arbitration, institutional arbitration, statutory arbitration, foreign arbitration, and domestic or international arbitration. The type of arbitration upon which the CAC focuses is domestic and international arbitration. In this case, arbitration is a figure whereby an impartial and specialized third party, called an arbitrator, acts as a judge in the process of resolving a dispute, handing down an arbitration decision which has the same legally binding effects as a court decision would (Cámara de Comercio de Bogotá, 2018b). Given that arbitration has a legal effect, the role of the arbitrator is necessarily given to a lawyer.

On the other hand, mediation is a negotiation between two or more parties facilitated by an agreed-upon third party. Skilled third-party mediators can lower the emotional temperature in a negotiation, foster more effective communication, help uncover less obvious interests, offer face-saving possibilities for movement, and suggest solutions that the parties might have overlooked (Lax & Sebenius, 2006). Mediation, is argued to lend itself particularly well to situations where the parties have become entrenched in their positions. From a practical perspective, it is a method of resolving workplace disputes that, it is further argued, seeks to avoid a more formal and often more confrontational route, such as grievance and discipline

procedures, and rather than attribute blame looks to rebuild damaged relationships for the future (Advisory, Conciliation and Arbitration Service (ACAS), 2011).

The concept of negotiation does not have a single definition, as different authors define it by different means. Fisher, Ury & Patton (1991), define negotiation as a back-and-forth communication designed to reach an agreement when two sides have some interests that are shared and others that are opposed. On the other hand, Thompson (2012) defines it as an interpersonal decision making process that is necessary whenever we cannot achieve our objectives single-handedly.

Finally, conciliation is a voluntary proceeding, where the parties involved are free to agree and attempt to resolve their dispute by conciliation. The process is flexible, allowing parties to define the time, structure and content of the conciliation proceedings. These proceedings are rarely public. They are interest-based, as the conciliator will when proposing a settlement, not only take into account the parties' legal positions, but also their; commercial, financial and / or personal interests (Dispute Resolution Hamburg , 2018). Like mediation, conciliation is a voluntary, flexible, confidential, and interest based process. The parties seek to reach an amicable dispute settlement with the assistance of the conciliator, who acts as a neutral third party. However, conciliation is a more formal process than mediation and it could generally involve the engagement of legal representatives, thus making it a more expensive process than mediation. There is, however, the added advantage that should no amicable solution be reached, the conciliator has the duty to attempt to persuade the differing parties to accept his own solution to the dispute (Law Reform Commission , 2010).

7.3 The Hermes Program

The Hermes Program is an initiative created by the Bogota Chamber of Commerce (BCC) with the purpose of transforming school conflict by means of ADRs and the prevention of bullying in schools in Bogota and the region. The program has been devised as a pedagogic alternative that provides tools for the development of social competencies in younger populations, although it is directed towards the entire school's community, not only including students but parents, teachers, and administrative staff of the schools as well.

7.3.1 Background

The development of the Hermes Program is preceded by two components. First, is a historical

background given that before the Hermes Program, other program addressing conflict in schools was created and implemented. Second, the legal background which created a need for the creation of a program directed towards addressing conflict in schools.

7.3.1.1 Legal Background

The legal background that set the foundations for the development of conflict resolution programs is based on a series of laws and articles established in the 1991 Colombian constitution and after it. The first one, corresponds to the article 116 in the 1991 Colombian constitution which states that individuals can be given the function of administering justice as juries, conciliators or arbitrators. On the other hand, there are the following laws: 23rd law of the 1991 Colombian constitution, the 446th law of 1998, and the 640th law of 2001. These laws stipulates by means of a series of articles, aspects related to conciliation in terms of proceedings, how to proceed according to the possible outcomes of the conciliatory process, which may be the effects of conciliation in equality, applicable norms to conciliatory processes, among others. These laws and articles are directed towards addressing issues related to conflict resolution by means of conciliation within a legal framework, however there are other norms, arrangements, and laws which are directed to addressing coexistence issues within the environment of the schools.

By means of the 4th agreement of 2000 of the Council of Bogota, the coexistence committees in public and private schools of the city were created. This meant that it became mandatory for every school in the city to have a coexistence committee. Furthermore, in 2013 by means of the 1620 law, the national coexistence network was created for tackling issues related to the exercise of human rights, sexual education, and the prevention and mitigation of school violence. Given the preceding laws, the creation of a program that tackled these issues in the schools' environment was imminent. The Hermes Program represented an opportunity to fulfill the educational needs related to the accomplishment of such laws and norms.

7.3.1.2 Historical Background

The Hermes Program background goes until 1997 when the BCC promoted a project known as “Ariadna – tejiendo los hilos de paz”, known in English as Ariadna – weaving the fabric of peace with the support of the Interamerican Development Bank. The project methodology and conceptual framework sought to approach interpersonal conflict between young people and to generate processes of transformation in order to change attitudes and responses (UNICEF, 2009). In order to do so, its target group comprised children aged between 11 and

18 who were the focus of the program.

In 2001, the Ariadna project was launched in 10 schools in poor areas of Bogota and its surrounding areas, it was later extended to 47 public schools, where it had 687 school conciliators, 582 youth representatives at the school conciliation serve, and covering 4,228 members of the school's community, including parents, management and administrative staff of the schools, teachers and students, who were informed of the availability of school conciliation services (UNICEF, 2009). Five years later, in 2006, the BCC and the Bogota Education Secretary signed an agreement for the expansion of the program and its integration into the district's public policy.

This model fostered the leadership capacity of students as school conciliators and promoted conciliation as an adequate response to a counterculture permeated by values of aggression, verbal and physical violence, as a way to solve conflict. In the course of developing the program those involved in it underwent a process of internal reflection and feedback from community dynamics. This process led to deeper analysis of conflict diversity as well as improving the understanding of the traditional modes of conflict resolution that were seen as legitimate in each community (UNICEF, 2009). This variation brought about the need for a new methodology that was better able to respond to the more specific characteristics of each group. As a result, the Hermes program for conflict management at school, named after the Greek god of negotiation emerged. This program also intended to include all the members of the school's community, however focusing in this case on emotional intelligence as a way to tackle conflict. The emotional intelligence perspective was also used to identify and develop the skills and competencies needed for conflict resolution.

7.3.2 Objective

The objective of the program is to offer members of the educational community a series of pedagogical tools with which to transform conflict. These works to reintroduce dialogue and agreement in an atmosphere of mutual respect where tolerance is a pervasive value.

However, according to Rodriguez (2014) the specific objectives that guide the Hermes program are stated as follows:

- Youngsters' personal transformation. Generate empowerment.
- Develop abilities and social competences, aside of teaching a technique.
- Generate a culture towards concertation and conflict conciliation.

- Get to raise more awareness about ADRs in the community without the need of having an advanced knowledge in law.
- Legitimate the other as a real other and treat him with respect. Generate interpersonal recognition.
- Produce a transformation in the concept of conflict.
- Generate spaces for peaceful coexistence.
- Develop activities in students and teachers to solve conflict through ADRs.
- Transcend beyond coexistence.

In order to achieve these objectives, the program is divided in different stages directed towards achieving different goals. This will be described in what follows.

7.3.3 Methodology

The methodology of the Hermes Program can be divided in 5 phases, the first 3 of which are related to the students training in conflict resolution techniques, whereas the other 2 are directed towards teaching and guiding students in conducting entrepreneur projects once they have been certified by the program.

The program intervention within a school, starts by conducting a diagnostic stage in which the contextual variables surrounding the school, as well as the school's composition, operation and conflictual variables are assessed by the consultants of the program. Once such diagnosis has been conducted, the consultants determine how the predefined structure of the program needs to be adapted according to the school's context and operation. After this has been conducted, the actual intervention begins.

The intervention of the program starts with the selection of a group of teachers who will be in charge of the operation of the program in the school. In order to select the teachers, the BCC conducts a meeting in which the program is presented to the school community, after this socialization, the program consultants ask for volunteers to participate in the program. Once the group of teachers that will be participating in the program has been conformed, it is time to choose the students who will be participating in it. Students' selection works in the same way as the teacher's selection. However, volunteer students need to belong to the grades over which the program is focused. This means that if the program aims at working with students belonging to 8th and 9th grade, the volunteers need to be in those grades.

Once the students have been chosen, their formation in conflict resolution begins. At first, the

students are only trained, and once they have reached certain amount of time within the program training, they start to participate in conciliatory sessions with other members of the school community who do not belong to the program in helping them to solve their disputes by working as conciliators. Such activities primarily take place during recess and lunch times in which the students set up conciliation tables supervised by teachers or BCC consultants.

It is expected for the students to be part of the training process in conflict resolution for around 2 years. After this period of time, and after fulfilling other requirements, they get certified by the BCC as conciliators and peace builders. This is why the BCC seeks to work with students who are at most in 9th grade, as they will at least be certified before graduating from high school.

Once the students have completed their formation in conflict resolution techniques and they have been certified, they move on to other two stages of the program. However, these stages are not related to conflict resolution but on developing entrepreneur abilities.

The Hermes Program work with the students with regards to conflict resolution practices does not end once they have graduated from high school. The BCC has designed a graduate network in which the former students receive benefits from the BCC while they help in the schools in which the program operates.

7.3.4 Graduate Network

In 2016, the Chamber of Commerce launched the Hermes Program Graduate Network which consisted in the formation and consolidation of a group of students who have been trained and certified by the program. The goal with which such group was conformed, was for the volunteers to work in the formation of new peace builders in their communities (Cámara de Comercio de Bogota, 2016). In order for them to join the network, they needed to be between 14 and 22 years old, to have been certified by the Hermes Program. As a counter consideration for their work with the community, they will receive advice by the Chamber of Commerce in developing social and productive entrepreneur ideas. Likewise, when possible they would be able to receive scholarships from the Fundación Universitaria Empresarial (Uniempresarial), which is an affiliated university of the Chamber of Commerce, in order for them to get access to additional professional education in business, management and entrepreneurship.

7.3.5 Operation of the Hermes Program

The following section is directed towards presenting some statistical information about the operation of the Hermes Program. The Hermes program not only operates in different localities of Bogota but also in several of its surrounding towns. However, given that this research focuses solely on Bogota, I will only present relevant information regarding this city and not its surrounding areas.

Figure 13. The Hermes Program in Bogota.

The city of Bogotá is organized in large districts called “localidades” or by its literal translation to English “locality”. Each of these localities is made up of several neighborhoods. Figure 19 (taken from Rodríguez (2014)) presents the localities over which the Hermes Program was operating until 2014. As can be seen, out of the 20 localities of Bogotá, the program operated in 17 of them. Until 2016, the program operated in 452 schools, out of which 319 were located in Bogotá (Cámara de Comercio de Bogotá, 2018c). Since 2007, 27.282.522 actors, including teachers, parents, students, and administrative staff of the schools, have been sensitized and trained by the program, whereas 317.504 conciliatory processes have been conducted, 8.202 conciliatory days have been organized, and 65 schools are currently self-managed with regards to the operation of the program. Self-management means that they no longer need the surveillance or assistance of the BCC personnel in operating the program or in organizing and setting the conciliation activities.

Since it began its operation, the Hermes Program has participated in several competitions, receiving multiple recognitions and awards. In 2008 and 2009, the program got the second place for best experiences in education for young populations. This award was granted in a Social Innovation Fair organized by the United Nations Economic Commission for Latin America and the Caribbean (ECLAC) and the W.K. Kellogg Foundation. In 2010, the ECLAC chose the program as one of the 25 most social innovative experiences in Latin America, out of which it was considered to be the best in school issues, and the one with the largest social development impact in the region. Finally, in 2011, the program participated in the World Chambers Competition organized by the International World Chamber in the category of “Best Corporate Social Responsibility Project” being chosen as one of the 5 finalists out of 71 projects from 41 countries.

Given that I have already introduced what the Hermes Program is, I will now focus on addressing how the Hermes program evaluation was conducted.

Chapter 8 – Hermes Program Evaluation

So far I have addressed the theoretical relationship between systems thinking concepts and evaluation, particularly on the use of boundary critique in Fourth Generation evaluation. In the following section I will focus on describing how the reframed methodological design of Fourth Generation evaluation was used to evaluate the Hermes Program.

In order to do so, I will review each one of the steps proposed in the reframed methodology in a more detailed manner. These steps are:

1. Establishing a relationship with the client.
2. Organizing to conduct the evaluation.

3. Identifying adequate stakeholders.
4. Reflecting upon the stakeholders' selection.
5. Developing, testing, and enlarging constructions.
6. Identifying and prioritizing relevant issues, claims, and concerns.
7. Reflecting upon the boundaries about "what" should be addressed by the evaluation.
8. Design of adequate tools for data collection.
9. Data Collection.
10. Analysis of Results.
11. Discussion and sharing of results.
12. Generating recommendations.
13. Producing reports.
14. Recycling.

Although these steps might be presented in a particular order, this does not necessarily mean that they could not be carried out in a different way or that an evaluation needs to be conducted in a linear way. Instead, this is done so that the reader might get a clearer idea of what was done with the Hermes Program evaluation and how each of its key issues were addressed.

When using the proposed methodology as well as the Fourth Generation evaluation original design, the evaluator should conduct the evaluation in a tailored way in which contextual elements in which the program and its evaluation are taken into account. On the other hand, one of the most important elements in giving a systemic character to an evaluation is the need to not only consider the interrelationships among its stakeholders and the issues it tackles, but also to find and understand these interrelationships in the processes that are being studied as well. Given these considerations, each one of the steps outlined above will be addressed more extensively for the Hermes Program evaluation in what follows.

8.1 Establishing a relationship with the client

The evaluation of the Hermes Program was conducted as the central aim of the developing of my PhD dissertation at Los Andes University for which the current document was written. For that reason, there are several considerations regarding the establishment of a relationship with the client and the role of the evaluator that need to be addressed and which might differ from the traditional way in which they will be managed when establishing a client-evaluator relationship in a more formal context. First, given my status as a student, a written contract

between the two parts (evaluator-client) did not exist as I was not intended to receive any economic retribution for conducting the evaluation. However, this does not have as direct implication the lack of a relationship. Instead, it follows a special status in which Los Andes University plays the role of an intermediate figure primarily in charge of concerting meetings, facilitating and arranging activities, working as a communication channel between the two parts (client-evaluator/student) and most importantly in giving credibility, trust, and a sense of responsibility to the work that was being conducted in the research.

Aside of establishing a relationship with the client, deciding the purpose with which the evaluation was going to take place was another important issue at this stage of the evaluation. According to Guba & Lincoln (1989), the purpose of the evaluation could be classified as formative merit, formative worth, summative merit, and summative worth (these concepts were explained in more detailed in previous chapters of this document). In that sense, the first decision regarding the purpose of the evaluation was to select between attributing a formative or a summative character to it. As the Hermes Program is an ongoing program which is aimed at continuing operating for several more years, the decision was directed towards using the evaluation with formative purposes. This means that its objective is to be used as a way of not only exploring the actual state of the program and its performance in specific aspects, but also by using the evaluation as a means of generating improvement in it so that changes could be seen in the way it operates and the way in which it impacts the members of the community toward which it is intended.

The second element upon which decisions needed to be made was selecting between focusing on merit or worth. According to Lincoln & Guba (1980), merit is context free, however worth can only be defined in relation to a particular context. This difference has a direct implication in the way evaluations are conducted when being oriented towards merit or worth. As merit is not context dependent, it could be assessed in different contexts in the same way without considering the factors which might be having certain influence in it. On the other hand, when assessing worth, it can only be done by achieving a deep knowledge of the local social, political, cultural, and value factors having certain impact in it. In that sense, our evaluation was necessarily worth oriented as one of its main traits was the need to take into account not only the contextual characteristics surrounding the program but also how the factors embedded within it need to be taken into consideration when conducting the evaluation. For example, the social characteristics involved in it have an impact not only in the way in which the program is designed and implemented but also on the way in which it is assessed and in

which the evaluation activities are designed as well.

Taking these into account, I proceeded to conduct the following stages of the evaluation.

8.2 Organizing to conduct the evaluation

One of the fundamental elements when organizing to conduct the evaluation is the need to understand the context in which the evaluand will be evaluated. However, this task should not only focus on comprehending them but also on addressing and incorporating them in one way or another when conducting the evaluation. In the particular case of the Hermes program, there is a specific and a general context which needs to be addressed. These contexts do not only affect the program and its evaluation, but are also fundamental elements to be considered in order to understand the foundations upon which the program was conceived and laid.

8.2.1 General context

The general context of the evaluation and the Hermes program is constituted by those elements of the national context of Colombia which have a direct relationship or implications for the program. Although theoretically it might seem easy to separate each of these aspects, in real life, most of them if not all are intertwined. For this reason a general description of what I consider as being the fundamental aspects and issues to be taken into account will be presented, so that the reader might get at least a general insight of how their interconnectedness takes place, as well as to familiarize him with the situation. Each of the issues will be presented in a separate manner, and at the end of the section they will all be related in a more structured way.

According to the Departamento Administrativo Nacional de Estadística (National Department of Statistics, 2017a), the Colombian population up to 2016 was around 48.747.708, from which almost 28%, that is 13.649.358, were considered as being poor. This type of poverty corresponds to what is known as income poverty which is defined as taking place when a family's income fails to meet a federally established threshold that differs across countries (UNESCO , 2017). Typically, income poverty is measured with respect to families and not the individual, and is adjusted for the number of persons in a family. Furthermore, poverty in rural areas is 3.2 times higher than that in urban areas (Moreno, 2014).

Aside of having this poverty levels, according to the Departamento Administrativo Nacional de Estadística (2017b), in 2016 Colombia was among the ten most unequal countries in the

world with a Gini coefficient of 0.517. The interpretation of this coefficient is given by an index of concentration of wealth which means that the higher the Gini coefficient, more unequal a country's income distribution is. These characteristics allow us to see that there is a large proportion of the Colombian population facing economic difficulties, having a direct impact in the quality of life and living conditions they are able to achieve.

On the other hand, Colombia has been facing a civil war for more than 50 years involving different armed forces. This war has partially ended with the signing of a peace agreement between one of the armed forces, the FARC, and the government in 2016. However, war is still existing with other armed forces such as the ELN, the FARC dissidences who did not stick to the peace agreement, drug cartels, and others. War has had a large impact throughout the social spheres of the country, particularly over the approximately 7.000.000 of people who have been displaced since it began. According to the United Nations Refugee Agency (UNHCR, 2016), Colombia has the largest displaced population in the world. However, when comparing the Colombian displacement with that of other countries such as Syria, which is the second country with the largest displaced population in the world, one finds some differences among them.

First, displacement in Colombia has been sustained for a long period of time as it is characterized as a low intensity asymmetric war. A low intensity conflict is defined as a

“Political-military confrontation between contending states or groups below conventional war and above the routine, peaceful competition among states. It frequently involves protracted struggles of competing principles and ideologies. Low intensity conflict ranges from subversion to the use of armed force. It is waged by a combination of means employing political, economic, informational, and military instruments. Low intensity conflicts are often localized, generally in the third world, but contain regional and global security implications (Army Air Force Center for Low Intensity Conflict, 1989)”.

On the other hand, its asymmetric character refers to the level of military disparity between belligerents (Geiß, 2006). Taking into account these definitions, the Colombian conflict takes a low intensity character as it involves different armed groups with competing ideologies, which is the reason for which the conflict emerged, as well as different types of conflict which include subversion, use of armed forces, and attacks to civil population. Colombian war has an impact on the security not only for the region but also globally as it is directly related with drug and arms trafficking in a global scale. When referring to its asymmetric character and taking into account the definition of the term, it acquires this character because it includes a professional army which is represented by the Colombian military forces, and

insurgent resistant groups such as the FARC (Fuerzas Armadas Revolucionarias de Colombia), the ELN (Ejército de Liberación Nacional) and the AUC (Autodefensas Unidas de Colombia). Having this type of war, means that large volumes of populations are not displaced at once (at least not by measuring it in terms of millions). Instead, displacement at a particular period of time affects hundreds or thousands, this means that displaced people is less visible and population movements are not instantly perceived but when they get accumulated in higher proportions through time.

Second, Colombian displacement takes places primarily at a national level, this means that most of its migration has an internal character from one place to another within the boundaries of the country and not to other countries. According to the data record of the Centro Nacional de Memoria Histórica (2015), until 2013 14% of the Colombian displaced population live in the most important cities of the country, 5% of which live in Bogota. This directly affects the living conditions of not only the ones arriving but also of the ones already living in the places to where this population arrives. All these elements, makes displacement a key contextual element for the country as it is an issue that has affected its entire population in a direct or indirect manner.

An additional element related to poverty, inequality and displacement issues which have been described so far, is the cultural and ethnic component. Colombia is a very diverse country not only for its wildlife but also in its population. According to the Departamento Administrativo Nacional de Estadística (DANE), in 2007, 10.62% of the population was considered to be afro-Colombian, comprising raizales, palenqueras, mulato, black, and afro descents, and 3.43% of the population belonged to any of the 87 recognized indigenous groups within the country, and 0.01% defined themselves as Romanian Gypsies. The rest of the population represented by an 85.94% considered themselves as not belonging to a particular ethnic group. However, further research conducted by the Federal Research Division of the United States in 2010, found out that 49% out of the 85.94% were mestizos (a mix of different races) and 37% were white.

When putting under the scope the displaced population and their ethnic belonging, Suescún & Jiménez (2011) found out that those who identified themselves as afro-Colombians, indigenous or Romanian gypsies were more affected by forced displacement than those who did not see themselves as belonging to a particular ethnic group. According to Suescún & Jiménez (2011), between 2001 and 2009 73.3% of the displaced individuals who considered

themselves as belonging to a particular ethnic group were afro-Colombians, 21.3% were indigenous, and 5.4% were Romanian gypsies. This means that not only most of the displaced citizens throughout this period of time belonged to a particular ethnic group, but it also shows that afro-Colombians are the most largely population affected by displacement in the country. As I had stated previously, as a result of being displaced, these populations move to the biggest cities of the country looking for more opportunities. However, the cultural and ethnic differences of those living in this places and the ones arriving, represent big challenges for everyone.

Another important variable to take into account is the unemployment rate. According to the Departamento Administrativo Nacional de Estadística (2018), for 2017 the unemployment rate of the country was 9.4%. This was the lowest unemployment rate recorded since 2001, when they began collecting the data. However, being employed is not a synonym of achieving certain level of quality of life, as for 2016 only 61.7% of households had enough income to minimally cover their expenses, whereas 23.1% did not have enough income to cover all their expenses (El Espectador, 2016).

Up to this point, the reader might have a glance about the general Colombian context and which of its elements are relevant for understanding and evaluating the Hermes Program. In the next section, I will address how these and other specific variables are of interest in a more specific context which is Bogota, the capital of Colombia, and the areas in which the program is currently running.

8.2.2 Specific context

Once the general context has been described, it is time to focus on the specific context surrounding the Hermes Program. The program, has been implemented in Bogota, the capital of Colombia, and some of its surrounding towns where the Bogota Chamber of Commerce has jurisdiction. This is an important issue as the program was developed and implemented by the Bogota Chamber of Commerce. The reason for this, is that the program can only use its monetary resources in areas under its coverage. This means that neither the program nor its money can be used or transferred to other Chambers of Commerce or geographical locations out of their control. The current evaluation was focus only on the schools in Bogota where the program was being implemented, for this reason the specific context I will be referring to in this document pertains exclusively to Bogota and not to its surrounding areas.

When addressing the general context, I described the dynamics of poverty, unemployment, inequality, displacement and cultural differences within the country. I will now address some of them as well as other relevant variables for the specific context surrounding the Hermes Program. In order to identify these specific variables, I asked the stakeholders of the program (students, teachers, the school's administrative staff, former students, parents, and consultants) by means of a workshop, which were the most relevant Political, Economic, Social, Technological, Environmental, and Legal variables surrounding the program. Their answers focused on the following variables: unemployment, economic deficiencies, drug sales, domestic and social violence, low education levels, low expectations and high dropout rates. Each of them will be addressed in what follows.

Bogota is the capital of Colombia, and up to 2016 its population was around 8.000.000 million inhabitants (DANE, 2016). From this 8 million, 817.000 were living below the line of income poverty (Observatorio de Desarrollo Económico, 2016), whereas 88% of its population, that is 7.040.000, was living in the levels 1, 2 or 3 of stratification. This means that taking into account the stratification levels in which the population lives, most of the population of the city could be considered as belonging to a middle-low to a low-class. This means that the majority of the population of the city lives in conditions which are close to border poverty or actually live in poverty conditions given their income and their capacity to fulfill their basic needs. On the other hand, as I have told in the previous section, Colombia has been in a constant war for more than 50 years, bringing as a result around 7 million of displaced citizens. According to the Victims Unit (Unidad de Víctimas, 2017), in the past 30 years, around 508.270 displaced individuals have arrived in Bogota as a result of conflict. On the other hand, the estimate of the Federal Research Division (2010) gave an estimate of 624.286 displaced individuals arriving in Bogota for the period between 1986 and 2006⁴. Although there is not an exact value of how many persons have arrived in Bogota as a result of displacement, the estimates show that it has been around half a million individuals.

Most of the displaced population arriving to Bogota locate themselves in the most marginalized areas of the city where living conditions are already difficult. It is in these areas where the Hermes Program primarily operates. For this reason, displacement, poverty and inequality are key variables for understanding the dynamics surrounding the program as

⁴ There is not an exact value on how many displaced individuals have arrived to Bogota because many of them might not have been accounted for as they need to register in a government data base and some of them do not do it.

well as the type of population it seeks to attend. The program works in the most vulnerable areas of the city which represent a mixture of economic and social difficulties in terms of poverty, gangs, drug trafficking and lack of job and education opportunities. Additionally, displacement represents a key issue as well not only because Bogota is the main hub for displaced citizens in the country, but also because it brings big challenges not only for the arriving population but also for the ones receiving them. Displacement could exacerbate feeling of resentment, anger, lack of forgiveness and hatred in individuals not only by growing in an environment which is constantly surrounded by insecurity and war but also by feelings of lost opportunities represented by those who arrive to compete for them in the same geographical space, which is the city.

In this regard, violence is another important issue. Throughout the first six months of 2017, 11.541 complaints regarding domestic violence (regarding couples not children) were filed in Bogota (Secretaria Distrital de Integración Social, 2017), where 99% of the cases involved women as a victim. On the other hand, the crime statistics of the city, are higher in some aspect than others. The rate of homicides of the city for 2017 was of 14 every 100.000, whereas the country's rate was of 23 every 100.000 (El Tiempo , 2017). This is an important statistic because it represented the lowest homicide rate in 32 years for the city. However, other crimes such as robbery are more common problems as according to a survey conducted by the Chamber of Commerce (2017a), 77% of those surveyed have been victims of robbery at least once. Aside of this, the results of the survey also showed that 50% of the surveyed considered that insecurity in the city had risen compared to the previous year.

According to El Tiempo (2018), the dropout rate in Bogota for 2017 was 2.1%. When students and teachers were being asked about the main reasons for which students left school, distance between the school and their home, economic difficulties at their homes, inadequate pedagogical techniques, child labor, education is not seen as important or necessary, lack of motivation, and school violence and coexistence difficulties were the most prominent answers (Secretaria de Educación del Distrito, 2017). The distance between the school and the students' home is an important problematic because it is related primarily with safety issues and drug dealing. During 2016, the local government identified 131 big places for drug dealing, known as "ollas", as well as 470 satellite sale sites close to public schools (El Tiempo, 2016). This is an important issue not only because the stakeholders have identified it as an important contextual factor surrounding the schools and the program (as all the schools of the program are public), but also because 4 out of 10 public school's students have tried

drugs at least once (El Tiempo, 2016). The presence of drug dealers, and drug distribution has an effect on the numbers of robberies committed in the schools' surrounding areas, the use of students for drug trafficking, drug consumption, violence, among other types of crime.

Once both, the general and specific contexts, have been addressed one of the main tasks of the evaluation takes place; that is the identification of the stakeholders.

8.3 Identifying the stakeholders

One of the key elements of Fourth Generation evaluation is the way in which it sets its ground in the dialogical/hermeneutic process in which the stakeholders are involved. However, in order to involve them in this process, they first need to be identified. I propose the identification and use of these stakeholders in two steps, the first one by a dialogical process between the client and the evaluation team (or evaluator) and the second one by involving the stakeholders identified in the first stage. The way in which both stages were conducted will be described in what follows.

Stage 1

The stage one of the stakeholder identification process has at its center a dialogical process between the client and the evaluator. In the case of the Hermes Program, this meant organizing a meeting between the Chamber of Commerce director of the program and the evaluation team, which in this case was represented by myself and Luis Pinzon (my thesis advisor). By means of a dialogical process between the three of us, we proceeded to identify not only the stakeholders but also which was the main interest of the Chamber of Commerce to be at the center of the evaluation.

The process of identifying the stakeholders was somehow simple as the three of us had a previous knowledge of the program and its general dynamics. For that reason, the identified stakeholders of stage 1 were:

- Students: students are at the core of the program as they are the one which the program intends to impact directly. However, when being conceived to be part of the research three different types of students were identified; students who have not been part of the program, students who were currently part of the program at any of the stages related to the training in conflict resolution techniques, and students who had been part of the program in the past and who had already been certified by it and who had not graduated from high school up to that moment.

- Teachers: the teachers considered to be relevant for the research were those ones who aside of being familiarized with it had received some training by the Chamber of Commerce and who were actively involved in its activities in their own schools.
- Parents: the case of the parents was very similar to that of the students as they could be divided among those who were parents of students who have been trained by the Chamber of Commerce but they themselves have not been part of any of its activities, those whose children had received training by the Chamber of Commerce and they themselves had received it as well, and those parents who had received some type of training by the Chamber of Commerce but whose children were not actively involved in it.
- Administrative staff: the administrative staff refers to the principal, the academic director and the coexistence director of each of the schools. Although the principals are the central decision making actor when deciding to implement or not the program in a school, in most of the cases after the program is implemented it is the coexistence director who is in charge of its operation (giving adequate times and spaces, handling the students and teachers etc.) and its surveillance. In this particular case secretaries and administrative staff of the type were not taken into account.
- Consultants: the consultants make reference to the Chamber of Commerce professional who are in direct contact with the members of the community and are the ones who develop the activities with them. They are given this name as it is the literal translation of the name of their role in Spanish. The role of the consultant is fundamental in the operation of the program as they are the pivot figure that connects the community with the management side of the program and they go back and forth between both sides.

Up to this point all of the stakeholders were identified jointly by both the evaluation team and the program director. However, in the dialogical process, the director of the program recognized an additional stakeholders who otherwise, would have been impossible to identify solely by the evaluation. The additional stakeholders was represented by the former students of the program.

- Former students of the program: in contrast with the students considered in the first group of stakeholders, the former students of the program had not only been certified by it, but at the point in which the research was conducted, they had already graduated from high school.

As a result of the first stage six different stakeholders were recognized, these were: students, teachers, parents, administrative staff of the schools, consultants, and former students of the program. These stakeholders were then used for conducting the second stage of the process of recognizing the relevant stakeholders for conducting the evaluation.

Stage 2

The second stage of the stakeholder recognition was developed along with other steps of the methodology by means of a workshop in which representatives of each of the recognized groups of stakeholders participated. Although the workshop was carried out with several purposes, a particular activity was directed towards reflecting about the boundaries of the stakeholders of the evaluation.

The input of this stage were the recognized stakeholders of stage 1 as they were the ones represented in the workshop. However, I wanted to express themselves how they understood the community of the program and the way the interrelationships in it took place. In order to do this, I conformed small groups with representatives of each of the stakeholders recognized in stage 1 they were asked to draw the community, later on when all the groups had finished, they were asked to socialize what they had drawn. The detailed results of this activity can be found in the results chapter of this document and its design can be found the appendixes as well.

In general terms, conducting this activity with the stakeholder represented a shift from the way in which we framed the adequate stakeholders in stage 1. The stakeholders' boundary of stage 1 ended up being narrower than the one considered by the stakeholders in stage 2. Most of the working groups, although not all of them, considered as relevant stakeholders the maintenance, cleaning, and janitorial services of their schools and the role they played in their daily life activities, as well as those related to the program. In theoretical terms, this reframed the original frame in which stakeholders were bounded as it did not represent a complete shift from the original conception but a widening process of it primarily because the stakeholders originally considered were still being considered by them.

8.4 Reflecting upon the stakeholders' boundary

As was shown in the previous step, the process of identifying stakeholders was carried out in two stages. The result of doing this was the establishment of two possible courses of action that could be followed, represented in two different ways of framing "who" should be taken

into account as a relevant stakeholder for the evaluation. Figure 20 and Figure 21 show the resulting boundaries of these processes.

Figure 14. Stage 1 Boundary.

Figure 20, shows how the client and the evaluation team by means of dialogue defined a boundary around the pertinent stakeholders to be involved in the evaluation by considering students, parents, teachers, the administrative staff of the schools, former students of the program, and consultants as being the relevant stakeholders. At the same time, other stakeholders although being identified were not seen as relevant. This boundary represents the stage 2 input where a reflective process around it would take place. In that sense, and by using tools described in the previous step, the stakeholders recognized in stage 1 were asked to illustrate (by using drawings) how they defined the structure of the community of the program, who were involved in it and most importantly which were the interrelationships among these stakeholders. The result of this process is shown in Figure 21.

Figure 151. Formed boundaries in stages 1 and 2.

Figure 21 shows how the stakeholders expanded the boundary considered by the client-

evaluator stage 1 process by not only identifying but also considered as pertinent the role of the janitors, cleaning and maintenance services of their schools, primarily arguing that these individuals had a direct impact on the daily life activities carried out in the schools' environment as well as being benefited, although not in a direct way, by the Hermes Program. Other possible members of the school's community were not considered as being relevant. One of the most important issues that concerned the client and the evaluation team was the impact gangs and drug sellers could be having in the schools. However, when conducting the workshop the stakeholders, primarily students, considered that this was not a big issue in the specific environment of their schools. For this reason, this type of stakeholders were not considered in none of the stages. Although in the workshop conducted with the stakeholders there were 7 different groups, there was consensus in most of them in considering the relevant stakeholders. Some of the groups kept the implicit boundary introduced in stage 1 whereas others expanded it by including the named new actors. It could be said that the stage 1 boundary had an implicit character because it was used to design the workshop and include the stakeholders recognized in it and not to impose it in the way in which the workshop was developed. For this reason the way in which it could bias the participants was reduced as much as possible.

At this point, reflection moved from defining different boundaries to actually choosing one to be used throughout the rest of the evaluation design and implementation. This decision making process is a primary responsibility of the evaluator or the evaluation team (if there is one) and it could be supported although not necessarily by the client. This is fundamentally because one of the most important factors influencing the selection of one boundary or another is the implications this might have on the implementation stage of the evaluation, the way in which the evaluation is focused, as well as the relevance given to the stakeholder by means of how its participation in the program is seen as significant or not.

In the particular case of the Hermes Program the decision of whether to keep the stage 1 or the stage 2 boundary lies on a series of questions: which would be the implications of involving janitors, maintenance, and cleaning services in the evaluation? In which ways could they contribute to assessing different issues of the program's operation and impact? How viable is it to include them in the different stages of the evaluation? The answer given to these questions determined the boundary selection. Although the role of the janitors and maintenance services is key in the local environment of each of the schools, considering them, as an active part of the current evaluation was not viable for several reasons. First,

involving them in the evaluation would represent a constraint in practice primarily because the possibility of reaching them to conduct the necessary activities for the evaluation implied a lot of time and bureaucratic processes which were out of the control of the evaluator and the Chamber of Commerce, this not only meant that the process of reaching them implied a lot of work but it also implied the need to consider if they were a sufficiently relevant stakeholder for the program to take the time to do it. Second, the effort of involving them was also related to not only their relevance within the conduction of the program but also on how much they could contribute in the evaluation of issues related to the program as they might have a general idea of it but have not received any type of training in the matter. Third, it was not really viable to include them in different stages of the process not only for the bureaucratic issues contacting them implied but also because as they were not directly related to the program, adequate spaces, time and incentives for their participation would not be given. For these reasons, I as evaluator decided to use the boundary depicted in Figure 19 comprising the participation of parents, the administrative staff of the schools, students, teachers, Chamber of Commerce consultants, and former students of the program.

8.5 Developing, testing and enlarging constructions

After deciding which individuals were going to be taken into account for conducting the evaluation of the Hermes Program, the selection of the issues, claims, and concerns to be tackled by it followed. In order to do so, I decided to organize a workshop with representatives of the different groups of the stakeholders so that different activities could be conducted to gather information about their perceptions regarding different issues, claims, and concerns related to the program. This workshop included 38 individuals between parents, teachers, former students of the program, current students of the program, Chamber of Commerce consultants, and administrative staff of the schools that were invited. The workshop was carried out in one of the locations of the Chamber of Commerce, however all the activities were designed and conducted by the evaluation team. The detailed results of each one of the activities conducted in this workshop can be found in the Results Chapter of this document. The current chapter intends to describe the general process with which each of the activities was carried out and the rationality behind them, for that reason there is a different chapter focused solely on describing results.

The development and design of the workshop was directed towards assessing the perceptions of its participants about issues related to:

1. Activities of the program: this refers to the adequacy of the time and space given for conducting the activities of the program as well as their perception of the adequacy of the activities conducted within the program.
2. Change: issues related to change do not necessarily need to be positive, they could be negative as well. Change could be directed to the internal dynamics of the school, home, or relationships among the stakeholders who participate in the program primarily directed towards the way in which they approach conflict, as this is the main aim of the Hermes Program.
3. Future change. As it is expected not everything within the program will be perfectly perceived for this reason I wanted to ask the participants of the program the things they would like to be different and how they would change them as well.

Why these issues?

There are several reasons why this particular group of issues were considered to be the main focus of the workshop. First, during the initial stages of the evaluation planning, the main interest of the Chamber of Commerce was that of assessing if the program was having results in the way in which its participant students were approaching conflict and solving conflictual situations when facing them in their daily life. However, in order to be able to gain a richer and wider insight about this issue, one should not only focus on the way in which the students approach conflict but also on the elements surrounding this approximation. In my perception this meant that the focus should not only be solely on the students but on the other members of the school community as well. This is because although change might be perceived in the student's behavior, change might also take place on the behavior of the other stakeholders as well both of which are significant for measuring the impact of the program. Second, there are several issues about the program operation which can affect its impact, for this reason I believe that asking the stakeholders their perception about what the activities and the way they related with the personnel of the Hermes Program was important for having a wider insight as well. In that sense what I wanted to do was to expand the focus of the evaluation by not only considering the interest of the program director but also by including significant things for other stakeholders as well. However, the point of doing this workshop was to find out which elements within this three areas of concerns were important for being further studied by means of the evaluation.

Taking this into account I structured the workshop and designed its activities by considering

the following constraints:

1. Limited time frame. The workshop could be conducted only once, this meant that I only had a single opportunity to gather and comprehend as much as I could regarding the perceptions of the stakeholders regarding the different elements of the program that I wanted to address. This meant that the activities needed to be designed in a way in which they could gather as much information as possible, and that the gathered information was not only extensive but useful as well.
2. Intensive session. As the workshop could only be conducted once, it had to be done in 8 hour time slot so that we had enough time to conduct all the activities. Intensive sessions pose a challenge because the activities need to be designed in way in which by conducting an 8 hours workshop, those participating in it did not get bored or easily tired.
3. Differences in age and levels of education among the participants. This constitutes the main constraint for the development of the workshop as the age of its participants comprised a spectrum of more than 30 years. On the other hand, not only age was an issue but also the educational differences between them, some of the participants were still school students, others had achieved only a school degree and others had professional diplomas. This meant that the activities did not only need to be creative so that they did not get bored, but also easy to understand and not too complex to implement.
4. Power issues. As is well known, differences among the roles within an organization pose difficulties in the way in which stakeholders relate. This was a concern given that for example, students in the presence of other members of the school community of “higher” degree such as teachers and administrative staff, might not feel comfortable to express their opinions. In that sense, it was key to provide an environment in which this issues could be reduced as much as possible. However, I was well aware that it is impossible to 100% remove this type of constraints from the situation. Power issues in negotiation will be tackled in more detail in further sections of this document.

Considering these constraints, I decided to conduct the workshop by applying 5 activities between individual and group activities. Each of these activities as well as the purpose with which they were designed will be described in what follows.

8.5.1 Questionnaire based on Critical Systems Heuristics

Critical Systems Heuristics (CSH) were designed by Werner Ulrich in 1983. However, in order to design the questionnaire I customized the questions by modifying their grammatical structure and the vocabulary they originally used. This was done not only because they needed to be written in terms related to the Hermes Program but also because the vocabulary used by Ulrich might have been too complex for the stakeholders to properly understand it. On the other hand, modifications were done by not considering exactly the same questions proposed by Ulrich but by modifying some and adding and deleting others. These modifications were oriented towards developing specific questionnaires for each group of stakeholders considered in the workshop as given the context of the program not all the questions could be asked to all the stakeholders and they could not be asked in the same way either.

The questionnaire was the first activity of the workshop and was intended to gather the perceptions of the stakeholders regarding different elements related to the program. Some of these elements included the contextual factors surrounding the program and their effects over it, the relationships among the stakeholders and how they had been affected by the program, perceptions of commitment from other stakeholders, key variables for the success of the program, personal motivations for voluntarily joining the Hermes Program, among other questions. The applied questionnaires can be found in Appendix A at the end of this document.

8.5.2 Crazy trains

As stated previously, one of the most important concerns related to the inclusion of different stakeholders in the workshop was the need to handle trust and power relationships among them. Aside of tackling these issues by the way in which the activities were designed, I decided to conduct a particular activity with the unique purpose of generating confidence among the participants of the workshop.

After the stakeholders solved the questionnaire based on the Critical Systems Heuristics questions, they were organized in small groups that included when possible at least one representative of each one of the stakeholder groups. From that point on, they were supposed to conduct the activities of the workshop in a joint manner. But in order to do that, they needed to trust each other as most of them did not know each other. For this reason, I decided to conduct the “crazy trains” activity.

Crazy trains is an activity designed to foster trust among groups, this activity can be applied in groups of children, teenagers and adults which is why I decided to use it. The general idea of the activity consists of gathering a group of people and organizing them in a line, every person in the line will be blind folded or with their eyes closed except the last person. This person, is the one in charge of guiding the rest of the group in a particular direction by touching specific points in the back of the person in front of him. The second person will touch the back of who's in front of her, this becomes an iterative process until the person in the front is reached. The goal of the activity was to generate trust by trusting in someone unknown to us with our senses, which in this case is our sight. An important element of conducting this activity is that it is very helpful for building trust among individuals independently from their positions as stakeholders within the organization and the program under study. In that sense, it does not matter if the team members are students, teachers or parents, their gender, knowledge, or age, as what is important is their ability to communicate by solely using their senses. This in a particular way mitigates the impact of power relations of which we are aware are present in the relationships among these stakeholders.

8.5.3 Drawing the community

After conducting the crazy trains activity, the next activity was focus towards addressing how the stakeholders conceived the structure of the community of the Hermes Program and the relationships among them. I thought that the best way in which this could be addressed was by drawing and explaining the drawing. I asked each one of the smaller groups to jointly, within each group, draw the community as they perceived it. Afterwards I asked them to explain their drawing and the relationships depicted in them. This activity had as main purposes; understanding how the stakeholders perceived the community, understanding the way in which the relationships among the members of the community were taking place, identifying differences among the conceptions between the different groups as well as with the perceptions of the Chamber of Commerce and mine as evaluator, identify which elements were part of a shared view of the program, and comprehend if the context surrounding the program and its community had any direct effect in it.

The participants were handed in a cardboard and markers so that they could make a drawing big enough to be shared and viewed by the members of the rest of the groups. I asked them to designate a single person to draw while the other members of the group as well as the one who was drawing discusses what they wanted to show with their drawing. The groups had 45

minutes for completing the activity. After all the groups had made their own drawing, each one of them was asked to share its meaning. At the end of each presentation, the audience (the other groups) had the chance to make comments, enrich the picture, and/or compare it with their own appreciation of the community.

8.5.4 Stories of Change

Basing my work in the Most Significant Change (MSC) Guide proposed by Davies & Dart (2005), I decided to ask the stakeholders to tell (if possible) their stories of most significant change related to the program. The main difference with what is proposed by MSC, is that in this case I did not only ask them to tell success stories but failure stories as well, as if real improvement in the Hermes Program is wanted, one cannot only stay with success stories but with those stories that imply the need to make changes in the program as well. The three basic steps of the process according to Davies & Dart (2005), consist of:

1. Deciding the type of stories that should be collected. In this case those stories had to do with the perceived impact of the program, whether it was negative or positive, by the members of each group.
2. Collecting the stories and determining which were the most significant. In our particular case, we did not want to decide which story was more significant but to let the participant stakeholders to tell their stories as they pleased. In that sense, if a stakeholder did not want to share its story with its group or the rest of the groups because it did not feel adequate, there was nothing wrong with that. On the other hand, if the members of the group viewed a story as more relevant than other, it was their choice but it was not a particular instruction given to them.
3. Sharing the stories and discussion of values with stakeholders and contributors so that learning happens about what is valued. This was a fundamental step of the process because by sharing the stories, other members of the community were able to realize different experiences that otherwise they would not have been aware of. At the same time, being able to share them gave them the possibility to identify similar or opposed experiences in their own lives. In that sense the entire process was enriched not only by what was shared but also by those who were in a “hearing” position.

The entire process of telling the stories, was based on the following questions: in the past six months, in your opinion which have been the most significant changes in the students’ behavior? Since the program started, how the relationships among the members of the

community have been affected? Among these changes which one do you think has been the most significant? Have you perceived any changes in the sphere of the school's community, aside of the ones related to the student's behavior and dynamics?

These questions were selected as the center of the discussion primarily because they satisfied the general interests upon which the entire workshop was designed. The first and second questions were directed towards addressing the initial purpose with which the evaluation was commissioned, that is measuring changes in the student's behavior. Giving a temporary location to the questions (6 months) gave everyone the possibility of framing the stories in a particular period of time which could be more easily related to the implementation of the program and not to other possible variables which might have affected the behavior and dynamics between the stakeholders. The third and fourth questions were directed towards other members of the community and not only the students. This is a fundamental elements as although the students are at the center of the program, most of the time its effects might not be seen directly in them but in the individuals surrounding them. After the stories were shared within each group, each group was asked to share their stories (if they wanted) with the other groups as well.

Each group was given 1 hour as well as a set of formats to be filled with in more detail about different elements of the stories. A sample of the format may be found in Appendix B.

8.5.5 Six Thinking Hats

This activity was based on the 6 Thinking Hats proposed by Edward De Bono (1985) which uses six different colored hats to express the thinking elements. The main reason for thinking with hats is that each one of the hats has a defined role given by its color which directs the attention for the role of the color and not to the person who is expressing its ideas. A huge benefit of using hats is that it is a very helpful way of changing the way in which an idea is seen. De Bono (1985) defines six hats with six different colors and functions:

- White hat: it is the hat of facts. It emphasizes facts, data and information and is characterized for being neutral and objective.
- Red hat: is the hat of emotions. Under the assumptions of this hat, one should never justify why a particular emotion is being felt. This hat includes emotions, forebodings, intuition, sensations and preferences.

- Black hat: it expresses negative judgments (this must not be confused with negative thoughts). It represents the identification of barriers, dangers and risks of an idea, fact or decision. Its goal is to point out failures in an objective manner.
- Yellow hat: it expresses positive thoughts and judgments about an idea, fact or decision. It always looks to future benefits, it must justify the optimism it represents, and always tries to get things done.
- Green hat: it is the hat in charge of new ideas and new ways of focusing things. It handles change and it ultimately represents creativity.
- Blue hat: it is in charge of managing the processes of thought and control. It plays the role of moderator between the other hats and it summarizes what has been said having a concluding role.

The thinking hats are very helpful for expressing new ideas and for assessing their viability by means of using each of its colors. For this reason I decided to conduct this activity by using the hats for proposing future changes in issues the stakeholders consider should be changed within the Hermes Program. Once again, by using this tool to conduct the activity, the power relations among the stakeholders could be tackled in some way as they do not represent their role within the school community nor their own self, instead each one of them represents a particular color and the ideas behind it.

In order to conduct the activity, I asked the stakeholders in each small group to play in rounds by assuming the different roles represented by each one of the hats. In that way, they would be able to present their ideas and critique in a positive way the ideas presented by others. At the end of the activity, each group was asked to socialize their ideas and the viability of them.

The instructions for conducting this activity included six hats of different colors handed in to each group, a format to be filled in for each played round and 1 hour to conduct the activity. The mentioned format could be found in Appendix C at the end of the document.

8.5.6 Final Debate and Sharing

After conducting all the activities of the workshop the participants were asked to share their general thoughts about anything they wanted while it was related to the workshop or the program.

After conducting all the activities directed towards making the constructions at an individual, small group and bog group levels, the next stage of the evaluation is the identification and

prioritization of the issues, claims, and concerns that appeared in the workshop. The detailed results of the workshop are presented in the Results Chapter.

8.6 Identifying and prioritizing relevant issues, claims, and concerns to be addressed

After conducting the workshop and analysis its results it was necessary to recognize and prioritize the issues, claims and concerns that would be directly tackled by means of the next phases of the evaluation. This meant that new boundaries needed to be drawn including the issues, claims, and concerns the stakeholders identified as being relevant for them and those which could actually be included in the evaluation, those which would be prioritized. In order to prioritize the issues, claims, and concerns a boundary critique process needed to take place unless the same initial boundary was going to be selected. The main reason for conducting a prioritization of issues, claims, and concerns to be tackled by means of the evaluation is that although a wide scope of the evaluation is desired, so that a more comprehensive approach could be achieved, widening too much the scope of the evaluation might become counterproductive as its results will become scattered through too many different areas without being able to give conclusive or consistent results.

Although I have already stated that there is an entire chapter focused on describing the workshop results as well as the surveys related to the data collection of the step 9 of the evaluation methodology, the main findings of the workshop, upon which the first boundary was built were:

- The structure and timings of the activities of the program need further assessment. The stakeholders perceived that they were not given adequate times and spaces to properly conduct the activities of the program. They also felt that some of the activities needed to be restructured to make them more attractive.
- Stakeholders' commitment was seen as a key issue.
- Although the program places at its core the student's training, the stakeholders highlighted the importance of placing the parents, the teachers and the administrative staff and principals in the center of the program as well.
- Having the adequate spaces and times to conducting the activities of the program was perceived as a big obstacle for teachers and students.
- Administrative requirements were perceived as having a negative influence in the quality of the impact of the Hermes Program.

- Incentive packages and dynamism in the activities and workshops of the program were repetitively mentioned by several of the stakeholders as a way of motivating students and teachers.
- The success of the program is built from the collaboration of everyone and not only the work of the consultants and the participation of the students.
- Wider Systemic Effects. Although for the Chamber of Commerce the main focus of the program is over the students, the workshop showed that its impact is wider than expected as it directly affects other stakeholders. Although this affectation is not directly seen by the Chamber of Commerce.
- Too many bureaucratic procedures hinder the effectiveness of the work of the Chamber of Commerce consultants.

Once a wide boundary was set by considering these issues, claims, and concerns of the stakeholders, reflection upon the necessity of drawing a narrower boundary for conducting the evaluation was the next step. In order to determine if a narrower boundary was needed, the following questions regarding the issues, claims, and concerns association with the central purpose with which the evaluation was established in the first place, their capacity of widening the initial boundary of the evaluation (step 1) without making it too wide, and providing insights over problematic situations that could be affecting the performance of those involved in the program and the activities carried out in it.

Taking these into account, I decided to reframe the boundary in a way in which several elements of concern for the stakeholders could be tackled. In order to do so, they needed to be reshaped so that they could be integrated in the evaluation process. The particular reason for doing this was the interconnectedness between some issues, claims, and concerns. If one of them was dismissed, it was very likely that an understanding of other issues, claims, and concerns could not be completely achieved. These issues, claims, and concerns could be summed up in the following general topics:

- Stakeholders' commitment with the program.
- Changes in the stakeholders' relationships as a result of the program.
- Adequacy of the physical conditions surrounding the implementation of the program.
- Frequency with which conflict takes place (in a period of time of one month).
- Behavior towards conflict.
- Success variables.

8.7 Reflecting upon the boundaries about “what” should be addressed by the evaluation

In this particular case carrying out a reflective process regarding what to include in the evaluation is conducted simultaneously to step 6. This is done as one needs to take each one of the issues, claims, and concerns identified by the stakeholders and put them under the scope of their usefulness for the evaluation. Although an issue may not seem relevant or useful now it might become useful later. For this reason, although a boundary reformation may lead to leaving aside certain elements, this does not mean that they cannot be addressed later. This could be done in further stages of the same evaluation or in future evaluations.

Reflecting upon the chosen boundary is a task that could be conducted by different stakeholders of the program and the evaluation team depending on the dynamics of the program, the constraints that limit such relationships (such as time, space, or budget), and the knowledge that need to be brought to the fore to do so. In this case, the task was carried out by the evaluator without the intervention of the stakeholders of the program, primarily because of time and budget constraints which limited the scope of the evaluation and the way in which it was design. Nonetheless, quality and evaluation standards should be a top priority.

8.8 Identification and design of adequate tools for data collection

The information collected by means of the workshop could be seen as an initial source of data for the collection of more specific data in further stages of the evaluation. The data collected by means of the workshop had a qualitative character for this reason and in order to fulfill the Chamber of Commerce requirements for having quantitative results as well, I decided to design a set of surveys directed towards to particular groups of stakeholders of the program; the students and the Chamber of Commerce consultants. The reasons for only conducting further data collection with a limited amount of stakeholders were primarily related to time, resources and safety issues.

Time. As this evaluation is the main subject of a research, the time available for conducting field work is very limited primarily because the main focus is not practical but theoretical. This has as a direct implication the need to focus more on how things need to get done than to focus on doing them. On the other hand, there are other activities related to conducting the research, not directly related with the Chamber of Commerce. For this reason, part of the available time is devoted in fulfilling them.

Resources. As I stated in the considerations related to time constraints, resources are also limited when the responsibility of conducting the evaluation relies in a single person. There

are not enough available resources to hire additional personnel for conducting field work. In practice, the relationship with the Chamber of Commerce also poses a constraint as doing so would be considered as inadequate. However, the help of the consultants of the Chamber of Commerce in applying surveys is very helpful given the practical constraints surrounding this evaluation.

Safety. Most of the schools that are part of the Hermes Program and which were chosen for being part of the study, are located in vulnerable areas with complicated social and safety issues. For these reasons, the Chamber of Commerce do not allow me as evaluator to attend the schools, as this could put my integrity at risk and they would not be able to take any responsibility in case that happens. For this reason, the Chamber of Commerce consultants are the ones who apply the surveys to the students.

For the previously described reasons, the survey stage of the evaluation focuses solely on students and consultants as they are the most reachable stakeholders.

The design of the surveys also needed to take into account the following constraints:

- Amount of surveys. In order to have a significant sample of students, the number of surveys needed to surpass at least 300 surveys (this will be explained later). For this reason, and given the limited resources for analyzing the data, surveys needed to follow a multiple choice structure without including any open question.
- Methods of survey application. The student survey was going to be applied by the program consultants. This posed a challenge because as there were too many persons to be surveyed and not too much available time to apply it, the survey could not be longer than a two faced page. This also represented a challenge because the consultants would not have any additional help in case the students had any doubt. For this reason we had a short meeting where I explained them how the survey was designed, how each question was supposed to be answered and how much time would take the students to answer it.
- Limited time frame to apply the survey. At the moment when the survey was supposed to be conducted, students were finishing their school year. This issue put a lot of pressure not only on me as evaluator but also on the consultants as aside of closing the working year with the students, they had to complete the survey as well in the same time frame which was already limited.

8.8.1 Students' survey

The final version of the student's survey included 12 multiple choice questions, some of which had a single answer and others that could have more than one answer. The survey format can be found in Appendix D. Other information asked to the students which is not included in these 12 questions, corresponds to their school, grade, age and gender. In general terms, the questions could be grouped in the following groups:

- Demographic information. This includes age, gender, age, school's name, grade, household composition.
- Adequacy of the given space and times for conducting the activities related to the program.
- Commitment. This refers to the perceived commitment from other stakeholders.
- Attitude towards conflict. This refers to how the student approaches conflict when being in a position where he is the victim, the abuser or an observer.
- The least and the most important aspects of the program for the stakeholders.
- Overall satisfaction.

8.8.2 Consultants' survey

The final version of the Consultant's survey included 20 questions. The main difference between the consultant's and the student's survey was that the former included both, multiple choice and open answered question but the latter did not. The two main reasons for this differences was the number of consultants, as they were less than 50, and the way in which they could be reached. As the consultants have a formal relationship with the Chamber of Commerce, it is easier to reach them than to reach other stakeholders of the school community. However, the biggest constraint for reaching consultants was related to being able to gather the 48 of them at the same time and same place. For this reason, we decided to apply their survey by using an internet based survey and not a paper based one.

The questions of the consultant's survey could be grouped in the following groups:

- Demographic questions. These are questions related to age, gender, level of education, field of studies, and time they have been working for the Hermes Program.
- Communication. These type of questions are related to the communication channels between the consultants and the stakeholders involved in the program.

- Commitment and factors that affect the commitment of stakeholders. The perception of the consultants regarding the commitment of the community stakeholders with the activities of the program, as well as which are the key factors influencing the stakeholder's commitment.
- Changes in the relations between stakeholders. Perceived changed, whether they are positive or negative, in the relationships among the stakeholders of the program.
- Success variables. In the consultant's which are the variables that affect the most the success of the Hermes Program.
- Consultant career development. This group of question are related to the future plans related to the consultants' career development and whether they have to do with the program or not.
- Overall satisfaction. Just as in the students' survey, I asked the consultants their level of satisfaction with the Hermes Program.

The reader may find the consultants' survey in Appendix E.

After designing the consultants and students surveys, data collection followed.

8.9 Data collection

Data collection was conducted by means of a process called Maximum Variation Sampling and not a Random Sampling. Conducting this type of sampling has as a direct implication is that it seeks to include a wide arrange of extremes. The principle that it follows is that if one deliberately try to interview a very different selection of people, their aggregate answers can be close to the whole population's. The way in which this was conducted was by including people from a wide spectrum of ages, being involved in the program by different means, and by playing different roles in the implementation of the program. This was primarily used for conducting the workshop. The student's survey instead was conducted following a random sampling as it included a lot of schools and students, as well as having a closed character by only including multiple choice questions.

Data collection for the students and the consultants was done in different ways and in separate periods of time. Both will be described in what follows.

In the case of students, the Chamber of Commerce consultants were the ones in charge of applying the surveys in some of the schools working with the program. As a result, 30 schools were part of the survey represented by 1059 surveyed students. All these schools

were located in Bogota and 17 consultants were part of the “collection team” who handed in the surveys to the students. All the surveys were paper based and it took the consultants around two weeks to complete the data collection.

In the consultants’ case, we used a google survey so that they could answer the questions whenever it was possible for each one of them. I gave them a period of 10 days to answer the survey. As a result, 42 out of the 48 consultants answered the survey.

8.10 Analysis of results

The analysis of results of the surveys was conducted by using a statistics package name STATA. In both cases, a descriptive statistical analysis including frequency histograms, variance, mean, range and other statistics. After conducting the descriptive analysis (for multiple choice questions), I proceeded to do hypothesis tests for questions with results that needed to be tested in a deeper way. For those open ended question which could not be analyzed by means of STATA, I proceeded to compute all the answers and to find repetitive patterns in them.

8.11 Sharing results and generating recommendations

Discussion, sharing of results and generating recommendations were two activities that were jointly developed given that in order to share and discuss results maximum variation sampling was used. Such activity shared similar characteristics to the workshop initially conducted with the stakeholders.

8.12 Producing reports

There are different ways in which reports can be produced with regards to the Hermes Program results. This documents constitutes a report on its own, however given its extension and the language in which it is written, further reports need to be written in order to fulfill the needs of the stakeholders of the program. For example, when writing a report for the client, namely the Chamber of Commerce, this report needs to fulfill demands such as being written in Spanish, describing in detail how every activity of the program took place, presenting statistical results and graphs as much as possible, and presenting a concise description of findings and recommendations. As can be seen, the Chamber of Commerce has a central interest in numeric results and statistical information that could be presented in graphs and charts. This is particularly because managerially such results are seen as more adequate to be presented in meetings as well as for discussing results.

8.13 Recycling

The current evaluation did not include any recycling of tools or results from previous evaluations. However, as I had previously worked with the Chamber of Commerce in conducting an evaluation, I was already familiarized with the program. This was very helpful particularly because this let me be aware of constraints that otherwise I would not have noticed. On the other hand, I expect the results of this evaluation to be used as sources of information for further evaluations or as issues to be tackled more deeply in further evaluations.

The following chapter will describe in detail the Hermes Program evaluation results.

Chapter 9 – Hermes Program Evaluation Results

In the following chapter I will present the results of the evaluation of the Hermes Program. The results will be presented according to the order in which the activities took place in real life. For this reason, I will first present the qualitative results of a workshop conducted with different stakeholders of the program. Thereafter, I will present the results of a series of surveys applied to two groups of stakeholders, namely the students and the program consultants. The student results only comprise quantitative results; for this reason this section will include histograms, tables, scatter-plots and hypothesis tests where needed. On the other hand, consultants' results include qualitative and quantitative results. In that matter, results aside of presenting the same type of graphs and tables as the students' surveys, will also include the qualitative results highlighting the most relevant information of the answers

given.

9.1 Workshop results

As explained in the previous chapter, the first activity that was carried out with the stakeholders of the program was a workshop aimed at recognizing and gathering information regarding the perception of different elements of the program such as the motivation of the participants, the perceived commitment with the program, not only of themselves but also of other members of the community, future desired changes in the program, positive and negative experiences with the program, among others. Five activities were carried out according to the desired goal. In the following section I will describe the results of each of the activities, however, the structure and goals with which each activity was conducted are explained in the previous chapter.

In total, 6 groups of stakeholders were identified and represented by 38 individuals in the workshop. These groups of stakeholders were identified taking into account the previous knowledge of the program, and also by the assistance of the program director who helped to recognize the importance of involving the former students of the program in its evaluation. These six groups of stakeholders were: current students of the program, former students of the program, school teachers, school principals and administrative staff, parents and grandparents, and program consultants (which are Chamber of Commerce employees and are the ones who have direct and permanent contact with the school members).

9.1.1 Activity 1: CSH based questions

The first activity that was carried out in the workshop was answering a questionnaire based on the CSH developed by Werner Ulrich (1986). The aim of conducting this activity was directed towards identifying the individual perceptions about different topics of each of the individuals that were taking part in the program. Although, the questionnaires were customized for each group of stakeholders, in general terms they followed the same structure and all the participants had to answer it individually. In Appendix A the reader will find the questionnaires that were handed to each group of stakeholders. As the questionnaires were written in Spanish, the reader will find the Spanish and the English version of each of the questions.

Results will be presented by taking into account the type of question asked, presenting the information in this way and not by stakeholder, will give the reader a more comprehensive and interconnected idea about the different responses.

9.1.1.1 Motivation

Each of the actors was asked about what motivated them to be part of the program.

Teachers. When being asked about what motivated the teachers to be part of the Hermes program, the most common reasons were related to the following statements:

- It is a program that helps prevent conflicts.
- Because it actively involves students.
- It develops leadership skills in students.
- It implies the active participation of the actors.
- It improves the conflictual problems of the schools.
- The multiplying role of the students with their families and community.
- Personal inspiration of their own children who are part of the program in other institution to implement the program in the school where the parent works.

As can be seen in the previous statements, the teachers' motivation was oriented towards the direct impact the program might have on the students in general and how they can work as a pivotal figure to generate a positive impact on other members of the community as well. This includes preventive and resolving elements as they seek to educate students so that conflicts can be avoided and adequate skills and behaviors can be reinforced, and they also seek to use the tools the program provides to solve actual situations that are taking place in the schools.

Administrative Staff and Principals. Their most frequent reasons to work with the program were:

- Learn how to manage conflictive students.
- Hooligans.
- The desire to learn new tools to work with students.

The administrative staff and principal's motivation differ from the teachers motivation in that the former focus primarily in problematic situations and students of their institutions whereas the latter do not.

Parents. Parents see the program as an easy way of solving problems and because of that their motivation is linked to a personal desire to improve not only themselves but also to help others and being able to improve their household relationships and dynamics. These is identified in the following statements:

- Multiply what they learn in the program (at home and beyond).
- Guide and help more people.
- Responsibility with the community.
- Responsibility with themselves, their children and their grandchildren.
- The name of the program was very attractive.
- The program is an easy way of solving problems.

Graduates. Some of the reasons of the former students of the program to join it were related towards their personal benefit and others towards the benefits of the community as a whole. Some of them expressed motivations regarding personal learning whereas others were more focused on the impact the program and their work might have on the community. The most common statements were:

- To serve as support.
- To change the way I see life.
- Because the scope of the program is directed towards the community as a whole.
- To improve the connivance in the institution.
- It is a program directed towards to existing problems (bullying).
- To make a difference.
- Empowerment.
- To learn and understand with the majority of the problems take place in the school.

Students. The student's motivation was directed towards:

- Adequately use of free time.
- To solve the school's problems.
- To inspire people to feel safe.

As can be seen, students did not focus only on themselves but also on the needs of others and how improving themselves could have an impact on other people's life.

Consultants. Different from how the question of the other actors was structured, the consultant's motivation question was directed towards how they perceived the participation of the students in the program and how they could guarantee a sustained motivation and commitment from them throughout the program.

Question: How has been the participation of the students in the program? How do

you think it is possible to guarantee a sustained commitment and motivation of the students with the program?

The following are some statements that represent the most frequent answer given by them:

- Students are the most motivated with the process. We work to create strong ties between the students and consultants, and between students and teachers.
- Students' motivation can be exploited more when the administrative staff and the principal of the institutions place the program in an important position. The benefits can be seen in the quality of the accompaniment given by the teachers.
- Motivation can be sustained through workshops and activities that are fun for students. Certification is voluntary but most of the time almost 100% of the students get it.

As can be seen through the statements of the consultants, in their opinion the students are the actors which are more motivated with the program and the process itself. Several of the consultants highlight the consolidation of the relationships between consultants and students and in further activities of the workshop (from this questionnaire was part of), the student also recognize the quality and the importance of the relationships they have with the consultants. Other important aspect is the structure of the activities designed for the students as they also determine the degree of commitment with which they involve in doing them.

9.1.1.2 Commitment

In this question each of the actors was asked to give their opinion about the commitment of some of the other actors of the community without considering themselves. Each of their opinion regarding the particular actor will be presented in what follows.

Teachers. They were asked about the administrative staff and principal, students, and parents' commitment with the program in the school they work at.

The teachers had different appreciations depending on their particular institutions, I will present some statements of what they answered:

- The students and teacher's commitment is not that good due to the fact that the program is just starting.

- The students willingness is perceived, however appropriation of the program from the motivated (the ones who were included in the program) parents, teachers, and administrative staff and principal is still needed.
- The administrative staff support has been good and useful (for adequate spaces and help with group dynamics). Certified students have a high level of commitment. Parents have not been involved in the program yet and they have not integrated voluntarily either.
- All the members of the community have been committed by it is still necessary to involve parents more. The parents from the certified students are the ones who have been involved the most.

In general terms, teachers consider that parents need to be increase their involvement in the program, they think that students have a very high commitment with the program, unless the program is on its initial stages of implementation which means that none of the actors have an adequate level of commitment, and in some cases the administrative staff and teachers commitment needs to be improved as well.

Administrative staff and principals. They were asked about the teachers, and students' commitment with the program.

- The students are motivated and they generate strategies. Teachers have the required commitment but they lack of time to participate in all the activities. The administrative staff do not trust in the students, however they allow them to participate in the proposed workshops.
- Many teachers do not get directly involved with the program due to time constraints. However, most of them are interested and use the Hermes tool. Students feel motivated about their role as peace promoters as well as being recognized and looked after by other members of the community due to the role they play.

A general aspect of the opinion of the administrative staff that attended the workshop was that teachers do not have enough time to commit 100% with the program, for this reason they cannot attend all the planned activities. Although they show a real interest in the program. On the other hand, students seem to be highly committed.

Parents. As parents play an external role from the school it is very difficult for them to assess the level of commitment of the different actors of the community. For this reason, they did

not have any question about commitment.

Graduates. Their question was oriented towards assessing the role each of the actors played in the execution of the program. As their former role was that of student, they were asked about teachers, parents, and the administrative staff and principal of the school. Some of the graduates' statements were:

- [Teachers] Gain trust with their children and with the students.
- Help and commitment are needed from many of them [all the actors].
- The support and the work of the teachers is the most important. Regarding the administrative staff, their work is related with the organization of conciliation meetings (time and spaces). Parents are the closest emotional motivation for their children, parents also motivate other parents.
- In some of the institutions the support of the parents, administrative staff, principal, and teachers is not felt. This makes it more difficult to motivate the students (more support from teachers, and administrative staff and principal is needed as they are the ones who spend more time in the schools).
- They [The adults] are the ones who generate a civic culture. The adults are the ones who motivate children to participate in this type of programs.
- It is necessary that all [the actors] work together to achieve a bigger impact.
- The administrative staff [not the principal] foster a good treatment between teachers and students.

The most important conclusion that can be drawn from these statements is that the commitment of the students is highly influenced by the commitment of the adults, primarily the commitment of the parents. If the students perceive a high commitment from the adults it is more likely that they will motivate themselves. On the other hand, the former students identify very specific roles for each of the actors within the program, as they attribute the success of the program as being constructed by all of them.

Students. They were asked about the level of commitment of the teachers, administrative staff and principal of their schools with the program. The following are their statements regarding the question:

- At the beginning there was almost no support from the teachers. As teachers and administrative staff starts to be included in the program, it is easier to promote it and to gain the commitment from the other actors.
- One of the members of the administrative staff and one teacher have been key for guiding students and understanding the problems.
- Physical spaces are given for conciliation, teachers and administrative staff are always willing to improve and support the program.

As consultants only have scheduled visits to the schools, teachers transform in a key element for an adequate development of the program. However, sometimes students perceive that they do not have enough credibility or support from other members of the school's community. At least the students that were part of the workshop, felt that they had the adequate spaces and times to conduct the activities of the program.

Consultants. They were asked about the commitment of the teachers and the administrative staff (including the principal) with the program, to which they answered:

- High level of commitment and co-responsibility from both, the teachers and the administrative staff. They are propositional, helpful and there is a lot of team work. They recognize the transformation it [the program] generates.
- The teachers and administrative staff commitment depends on the empathy that the consultant generates in the school. When the commitment of the teacher is low, it is necessary that the consultants generates strategies and spaces to know the reasons why this is happening in order to promote actions to improve the situation.
- The program counts with the commitment of three to four teachers and maybe one member of the administrative staff that show interest in offering solutions to improve the quality of life of the students, so that they can acquire abilities and tools to solve daily life problems. The majority of the teachers and administrative staff see the program as a burden, a commitment that does not add to their personal interests. The program is perceived as an opportunity to receive a certification, however the actors do not have enough incentives to participate in something that does not have economic benefits in a medium or long term.
- One should be very strategical to achieve the objectives of the Chamber of Commerce in many schools where the institutional commitment is low.

- In many cases the importance of the program is expressed but they are not transformed in facts. Teachers and administrative staff commitment is relative, as sometimes one can find apathetic people and in other, very collaborative one.

In general terms, consultants perceive the commitment of the teachers and the administrative staff of the schools as dependent of a series of variables, namely the relationship among the school's personnel and the consultants, the availability of time with which they count, whether if the teachers perceive the program as useful or not, and if they recognize any benefit from committing to working with it. For this reason, the commitment of the personnel of the institutions is highly variable. As can be seen in the different statements, some of the consultants perceived the teachers and the administrative staff commitment as being very high, whereas others perceived having a high level of commitment from some teachers and not from others, and in other cases consultants felt that the teachers' commitment was not good enough.

9.1.1.3 Scope

The actors were asked about the appropriateness of the scope of the program, taking into account the social, economic, political, and cultural characteristics surrounding the program and the places where it is being implemented.

Teachers.

- It is appropriate, however it has not been implemented in the institution for too long (2 to 3 months).
- The scope of the program is adequate because it works towards building the country and it teaches leaders capable of appropriating from the reality of the country.
- Its scope is adequate but it should integrate pedagogic strategies for topics that produce reiterative behaviors.
- Its scope is adequate because it could let to improve the connivance manual of the institution.

In general terms, teachers believe the program has a good scope because it tackles relevant problematic situations for the educational community and the reality of a country like Colombia.

Administrative Staff and Principal.

- Times and group size are too narrow to generate more conscience.
- The improvement should be in supporting elementary school more.

The administrative staff who participated in the program believed that although the structure of the program might be adequate it should widen the number of students and teachers that are included in the workshops as well as focusing more in younger children in elementary school and not in high school. In this way, it would be easier to prevent conflict rather than solving it.

Parents.

- It is well focused but I would like it to include fieldtrips that include parents.
- It could include talks about forgiveness.
- Include talks about health and also the parent's help in the meetings.
- Parents and teachers should have more leadership in the program.
- Work with displaced population.

The parents themselves believe that their role in the program should be more central to the development of the program. They would also like to include topics such as forgiveness in the topics of the program and to include marginalized population in it as well.

Graduates. Just like the parents, graduated students would like to widen the limits of the program but in their particular case, their interest is directed towards working with conflictive students, in this case this would not be directed towards preventing conflict but to teach them how to solve it. On the other hand, former students also consider that the activities of the program should be more dynamic so that they can catch and keep the student's attention. Some of their comments were:

- Widen the scope of the program by working individually with conflictive students.
- Tackle topics such as family conflict resolution as it will be tackling the root of many problems the students have.
- Turn the program into something more dynamic.
- More activities that will allow the students to know the figure of Hermes, so that the program will be more recognized.
- Working with more steadiness with schools with the most difficult cases.

Students.

- Efforts should be made so that schools implement the program in a more solid way.

Regarding the scope of the program, students would like it to be stronger in its implementation. By saying this students mean that they would like the different stakeholders to take it more serious, as this will facilitate the development of the different activities and workshops of the program.

Consultants. They did not have questions regarding this topic.

9.1.1.4 PESTEL Variables

Through this question, the actors were asked about the political, economic, social, technological, environmental, and legal variables that have a direct impact in the implementation of the program.

Teachers.

- Large economic deficiencies, large sale of psychoactive substances, high levels of intrafamily and social violence, low educational levels, and high drop-out rates.
- Most of the people are displaced, families are very dysfunctional (lack of values and vigilance), poverty and high levels of unemployment.
- Families belong to 2 or 3 levels of stratification, they do not work as employees of some company, and instead they work in their own businesses. There are no cultural or intelectual expectations.

Administrative Staff and Principal.

Their answers were inconclusive as it was very difficult to understand them.

Parents.

- Consider the economic difficulties of a large proportion of the community.
- Authoritarian pedagogical models and sexist social models.
- High levels of domestic violence.

Graduates.

- With the peace agreement people realize that peace goes beyond the classroom and the community and that it must be built by everyone.
- The monetary capital of the program is being wasted due to the lack of commitment of some of the actors of the program.

- It is necessary to work in the cultural appropriation of the community and the locality⁵.

Students.

The students question was directed towards asking them about the presence of gangs in their school or its surroundings that limited their capability of working with the program. All the students answered negatively to the question, this means that there are no gangs in the schools of the students that were participating in the workshop, or at least the gangs did not affect their performance within the program.

Consultants.

- The possibility of the arrival of demobilized combatants or people that will arrive to the institutions due to the signing of the peace agreement.
- The educational policies that should support with more strength proposals like the Hermes Program.
- The teachers' salary is low, making their motivation to be low.
- The program generates a social transformation that affects all the areas of the community (critical thinking, reflexivity, inclusion and active participation). The administrative staff interests negatively affect the students and teachers training as they do not stimulate and instead make difficult the implementation of the program.
- At an economical level the commercial value of the program should be evaluated, taking into account that this is at the core of the goals of the Chamber of Commerce.
- The cultural and social factors are positive as they contribute towards the transformation of the communities that have lived a sustained conflict. There is a high diversity in public schools (among them and between them as well), for this reason the program has to adapt to the specific characteristics of each of the institutions through diagnostic phases.
- Social characteristics: floating displaced population, high levels of unemployment, low socioeconomic levels of stratification (1 and 2), and educational level of the parents is low, low level of accompaniment from the parents.
- Students seek to be recognized for their participation in the program.

In general terms, the different actors recognize common characteristics in the schools were

⁵ A locality corresponds to the district in which the neighborhood is located.

the program is operating. These are related to large economic deficiencies, displacement, difficult household relationships, and low expectations regarding the future of the children. This last variable refers to the fact that as the parents did not have many educational opportunities, social mobility, and economic stability when growing up, they do not expect their children to have them either. Displacement is due to the internal conflict that has been affecting the country for the past 50 years which is now coming to an end (at least with one of the subversive armed groups). Most of the displaced population of the country has moved to the capital and for this reason the schools are filled with students who have experienced displacement as a result of the internal war.

Regarding the particular contextual characteristics surrounding the program, the program needs to have more support coming from the administrative staff of the schools, as a real commitment coming from them has an impact on the teachers and students commitment as well, and from the development of educational policies that could help strengthen the establishment of the program. On the other hand, recognition is an important element that would boost the student's commitment.

9.1.1.5 Impact of the program

The stakeholders were asked if they believed the program has had an impact at a classroom, school or community level and how they could perceive that impact.

Teachers.

- The program is very helpful in the classroom.
- In order to consolidate the program, it is necessary to increase the appropriation of the different stakeholders of the program.
- More incentives from the administrative staff and the principal of the schools are needed to get the parents to involve more in the program.

Administrative Staff and Principal.

- The program has generated changes in the conflicts at a group level, it has generated ties among classmates, and it has trained propositional leaders.
- The students that participate in the program are recognized as a positive pathway for conflict resolution.

- The trained youngsters have been transformed by the program. They have stopped being part of the problem and became part of the solution.

Parents.

- I have implemented what I learned with other parents.
- It is necessary to direct the impact towards a structural level.

Graduates.

- It [the program] makes people reflect.
- It [the program] has changed the way of seeing life and it has gained the interest of the community to participate in it.
- It generates social consciousness among children that is later transferred to adults (parents).

Students.

- Through sensitization, students have been made more aware before performing wrong acts.
- Physical assaults decreased and students turn to the administrative staff more often. Coexistence in school has improved.

Consultants.

- The principal are more aware and committed, whereas the teachers are more passionate.
- To mitigate "Type V" conflicting situations.
- Recognition of solutions to the conflict based in human relations.
- There are institutions where the training in human values is left aside and the academic component is placed in first place.
- To supervise, advice, train, and support the connivance committees, demands an improvement in the profile of the consultancy.
- Leadership of teachers and students belonging to the program, alliances with the connivance committee, institutional recognition of alternative methods of conflict resolution, inclusion of the program in the connivance manuals.
- Group transformation, conflict resolution through negotiation tables, improvement in the climate of the class.

Stakeholders recognize the impact of the program at different levels. Teachers, administrative staff, and students recognize an improvement in the classroom and in the students' behavior. Parents recognize an impact of the program in their own behavior and how they have apply it in wider spheres that are not related to their household or to the school. On the other hand, consultants also recognize the importance of increasing the commitment of the different stakeholders in consolidating the program, as well as an improvement perception not only in the classroom but in the consolidation of the connivance manual and the recognition of different ways of solving conflicts.

9.1.1.6 Success variables

The actors were asked about the variables they believed affect the most the success of the program. The following are the ones mentioned be each of them.

Teachers.

- Lack of time and commitment from the school, the impossibility to prevent "Type III" conflict as it keeps taking place even with the presence of the program, the students and teachers' commitment has been really good.

Parents.

- Trust, respect, self-esteem, commitment for others and responsibility.
- The spaces, the professionals of the Chamber of Commerce, and the information that is given to the community to understand the Project.
- Motivate the parents so that they will get involved in the education of their children (particularly in the program).
- One of the main problems is violence, the program should tackle this. It is necessary that the program becomes an institutional culture and not just "posing".
- The support of the Administrative Staff, the professionalism of the consultant generates trust, the questions that appear in the act in a concise way and the students' interest for the program.

Graduates.

- Punctuality, good attitude, communicating schedules and timetables of the activities, tolerance and respect.

- Training, motivation workshops, and strengthening the role of the consultant as this is something that creates ties of appreciation towards them.
- Make the program more dynamic and change the routine a little bit, achieving more freedom for the students so they can talk with the consultants and the graduates of the program.
- Active participation and collaboration from all the actors, having the adequate spaces to develop the activities of the program, transmit a clear message about the program.
- Support from the schools.

Consultants.

- Clarity in the comprehension of the program, a genuine concern for the wellbeing of the other. Team spirit, spend time making the activities powerful, teachers and administrative staff that promote the program, teachers who are aware of being a role model, passionate consultants, and a genuine concern for the training and learning of the other.
- Disposition and willingness of the administrative staff, as they are the ones who decided what program to execute.
- Review administrative processes to reduce the pressure regarding goals and indicators.
- The consultants profile, the institutional motivation (the institution should be willing to give the required spaces for workshops and activities), the principal's support, it must exist a group of interested teachers, administrative staff that include the program in the institutional Schedule, and students willing to get training.
- Co-responsibility, credibility in the program, tools given by the Chamber of Commerce.

Through the opinions of the different stakeholders, one can recognize repetitive patterns among what they identify as success variables for the program. These variables are related to the structure of the program, the commitment of the different stakeholders, and the values and moral principles of the participants. Regarding the structure of the program, the stakeholders argue that it must be more dynamic so that it can attract and keep the interest of the different stakeholders, it should be innovative and creative with the proposed activities and workshops, and most important of all it should be able to have a real impact and an adequate operation, so that it really works and not only poses as one of the stakeholders said.

When referring to the commitment of the stakeholders, it is important to note that in previous questions different stakeholders also spoke about the importance of having the support of the administrative staff of the schools as they are the one encourage the students to participate in the program, not only by motivating them but also by giving them the adequate times and spaces to develop the activities of the program. Achieving a high commitment from the parents also affects the commitment of the students as they see them as supporting figures in their participation in the program. By getting an appropriate level of commitment from the different stakeholders of the program, it is possible to establish a shared responsibility among them. This will turn into gaining more trust and credibility for the implementation of the program (along with strengthening its activities as well). Most of the stakeholders highlight moral traits that should be found in the stakeholders that will positively affect the implementation and the success of the program.

9.1.1.7 Other actors' perception about impact and improvement

Actors were asked about how they believed other members of the community perceived the improvements the program is producing, how does these perspectives articulate and which one is the dominant one among them.

Teachers.

- High expectations inside the school.
- It is important to actively involve the community [in the program].
- It is hard to get an appreciation from all the members of the community. For this reason, many teachers do not get involved in the program and others do not perceive an efficient performance of it, which is why they do not participate.
- In general terms, the program is perceived as positive but some teachers have trouble in allowing students to participate in the activities of the program.

Administrative Staff and Principal.

- The project is just starting to be implemented in the school, which is why people do not really know about it. An open day could be helpful for the students that belong to the Hermes Program to be known by the community.
- The program is part of the connivance manual, in that way it gets direct feedback from different members of the community as it has a clear goal which it aims to achieve.

Parents. The parents question was directed towards expressing their own personal opinion about the program and how it is being implemented in the school their children attend.

- The program has enabled students, teachers, and parents to be trained.
- It [the program] has taught parents to listen, interact and share with each other. The school is interested in the welfare of children.
- Would like the program to include more parents.
- The consultants are seen as well trained and assertive. The program includes pertinent subjects within reach of the participants. It still needs to get promoted more and let other know the work that has been done in the past years. It is necessary to strengthen the participation of teachers and parents.
- The schools is very helpful with the implementation of the program and it has taught the mother how to be more tolerant with her children.

Graduates. They did not have questions about this topic.

Students. They were asked about how other students and teachers which are not part of the program perceive it, as well as how do they feel about it, and which were the things they liked the most and the least about the program.

Regarding other people opinion:

- The program is respected and the teachers (even if they do not belong to the program) are very supportive and they make the students feel good as they congratulate them for their work and compromise.
- Other people were "delighted" with the program and what others say may be used to strengthen and improve it.
- A group of people perceive it as an intelligent and functional initiative, others believe that it is not strong enough to solve and avoid conflicts, whereas others see it as an alibi so that students can miss classes.

What students like the most and the least about the program:

- The conciliation activities and the workshops conducted by the Chamber of Commerce.
- Having support to reach goals. The interaction with other kids is funny and pleasant.

- See the smile in the faces of those who solved their conflicts with the program. It is very annoying that students who do not have problems still go to conciliation sessions; this is a waste of time.

Consultants. They did not have a question on this topic.

Although other members of the school community perceive the program as something positive, a general perception among the stakeholders is that the program still needs to be socialized and advertised more through the different spheres of the school community. By doing so, it will be possible to involve more parents, teachers and students in its implementation.

On the other hand, the aspects the students like the most of the program are primarily related to the execution of conciliations and the results they achieve with them. However, one of the students expressed that what he liked the least of the program was wasting his time with students who went to the conciliation sessions to make him waste their time because they did not have any problem to be solved.

9.1.1.8 Self-Reflection

In this set of questions each of the actors was asked about their personal commitment with the program as well as about the challenges they have faced with their participation in it.

Teachers.

- Motivate students and be an active part of the program.
- Commitment to help the students.
- Propose the connivance project derived from the program.

Challenges:

- The refusal of some teachers to give times and spaces for training.
- Convince the administrative staff on implementing the program in the institution.
- Fear of not being able of reaching the students.

Administrative Staff and Principal.

- The biggest challenge was the transformation of students into conciliators and project leaders.
- Work is under way on how the principals and administrative staff trust or have confidence in the activities of the Hermes group.

Challenges:

- The priority challenge has been to articulate the Hermes program as a requirement or step of due process in the connivance Manual. As well as raising awareness among some teachers to give strong support to the program.

Graduates. They did not have questions regarding this subject.

Students.

- Students have been important in the planning, implementation, and presentation of results of workshops in front of parents.

Challenges:

- It poses as a great challenge to give both time and dedication to sowing and conciliation evenings.
- Get people to agree and understand which are the real problems they have. It is expected that from participation in the program, people will trust the students more.

Consultants.

- Multiplier, generator of behaviors that are oriented towards inspiring others. Motivator, listener, and support.

Challenges:

- To fight and to contribute to remove the feeling of hopelessness of the teachers. Understand the logic with which administrative staff and principals operate, to build real action plans.
- Create processes of act awareness that can be achieved in each of the teachers.
- Safety issues in specific contexts.

- Generate strategies so that parents make an adequate accompaniment.
- Do not give up taking into account the difficulties in the community. Build relationships with the actors
- A sustained support and accompaniment sometimes is very difficult.

All the stakeholders who were asked to make a personal reflection about their commitment with the program had a clear identified role in the program, whether it was related to implementing the program, sharing results or planning the activities. On the other hand, challenges come from different sources according the particular actors. In the case of the administrative staff and the teachers, their challenges were related to convincing the teachers and the administrative staff of the worthiness and importance of the program. Trusting in the program and its activities has an effect on having the adequate spaces and times to conduct the activities of the program, and to having the possibility of articulating the program with the connivance manual of the schools.

On the other hand, the students' challenges are related to having enough time to spend in the activities of the program and to help people recognize their problems. Finally, in the case of the consultants, they present an extensive set of challenges related with almost all the stakeholders of the program. The transversal challenges are related with achieving an adequate commitment from all the stakeholders, building strong relationships with the different actors, and not giving up despite the difficulties surrounding the program and the schools they work with. The particular challenges concerning the teachers, are related to getting to realize the importance of being conscious of their acts as much as they do it with the students, and to remove the feeling of hopelessness they might feel due to the particular difficulties of the contexts they work in and also from their personal lives. Challenges related to the administrative staff are oriented towards comprehending the logic with which they work. Once the consultants have achieved this it would be easier to identify and use the adequate strategies to get them to be involved in the program. Finally, the challenges they pose regarding the parents, are relating to generating the adequate strategies for increasing the accompaniment of their children and their interest in the program.

9.1.1.9 Beneficiaries

Actors were asked about who they believed were being benefited by the program, and whether they considered them as being the adequate beneficiaries, or if by the contrary the real beneficiaries should be others, or if it more beneficiaries should be included.

Parents.

- The beneficiaries are the students and families of those parents who have been trained.
- Parents, students, families and teachers, although there would be other sectors of society that could be linked to the program.
- Everyone in the school is benefited, however from 2000 children in the school only 4 parents are linked to the program.
- A larger population should be included in the program (parents, teachers and administrative staff).
- More fellowship has been seen among students.

Consultants.

- Students are directly benefited and indirectly it benefits tutors, teachers, parents, students of the educational institution and administrative staff. I would invite the community surrounding the school (shops, supermarkets, video games) and it would be helpful to include the support of the Ministry of Education with a public policy.
- Strategies to include security staff and grooming ladies as they are also part of the community should be developed.
- They benefit parents, students, teachers, administrative staff, and other educational community in general. Other actors such as neighbors are permeated as well.
- Students receiving the training directly (direct beneficiaries). Students who use conciliations, educational community in general (indirect beneficiaries) through workshops to teachers and parents and activities to raise connivance awareness.
- The actions of the program could be carried out with all the teachers, carrying out more in-depth training so that connivance is taken as being the responsibility of all and not only of those who are part of the program.
- It [the program] benefits the community in general (students, teachers, managers, and parents).

Consultants find the program as benefiting the community as a whole (students, teachers, administrative staff, and parents). However, the parents do not recognize the current impact of the program as being that wide. In the parents opinion although the program is having an impact on the students, the degree with which it is impacting the other spheres of the community is too narrow, primarily because there are not too many parents participating in the program. Regarding this matter, consultants and parents believe that the program's

training should be deeper and extended to other actors.

The rest of the actors did not have this question.

9.1.1.10 Training

Teachers.

- There should be more incentives for students (the Chamber of Commerce used to give vests).
- There is a great commitment from the Chamber of Commerce.
- They [the Chamber of Commerce] provide key elements for the application and accompaniment of the program in the school, it provides tools that can be applied on a regular basis in the dynamics of the institution and in the use of the connivance manual, due process, and conflict management.
- It makes contributions for the revision of the connivance manual.

Administrative Staff and Principal.

- The project leader has generated motivation and has allowed a real integration of the group.
- The graduates have made an accompaniment to the students of the institution and this makes it easier to know other conciliators.
- The staff of the Chamber of Commerce is seen as ideal, responsible, affective and motivated. They reflect values such as honesty, responsibility and friendship.

Graduates.

Graduates were asked about the usefulness of the skills they learned with the program and how they can put them in practice.

- I can apply the skills learned in everyday life. These skills are: the ability to speak, listen, be patient, considerate, give advice, and always have a good attitude.
- Skills: be tolerant, always seek a solution to any problem, learn from the situations attended in the conciliations, identify real problems, appreciate goof connivance, be patient, keep in mind not breaking the limits of coexistence, dialogue, ask for forgiveness, and forgive.
- Handle a better vocabulary, learn to listen, have patience, and control.

- Patience, leadership and listening skills.
- Learn to listen others, think before acting, and understand that every wrong doing has a bad consequence.

Teachers and administrative staff were asked about their opinion of the training given by the Chamber of Commerce, whereas the graduates were asked about what they had learned with the program and how they have used it in their lives. The former two, had a very positive opinion not only about the training they and their institutions were receiving but of the consultants as well, attributing personal values such as honesty, responsibility and friendship to them. The administrative staff representatives also recognized the work of the graduates in accompanying their schools, and the role of the project leader in motivating those who are participating in the program. On the other hand, teachers also highlighted the importance of having more incentives for the students.

Among the skills graduates have learned through the program, are: the ability to listen, speak, be patient, leadership, thinking before acting, being more tolerant, dialogue, and forgiveness.

9.1.1.11 Additional Comments

Teachers.

- Teachers should become more involved in the program so that it can be exploited more.
- That Los Andes University does these same workshops with the students, parents and teachers involved in the program.
- The Chamber of Commerce lend or give away the vests that differentiate the students so that they can be identified.
- Create a Peace Laboratory for families in the schools with people such as psychologists, lawyers, and counselors who function as guides in the solution of cases.
- More time for workshops and more work from the leaders of the Chamber of Commerce is needed.
- Bring the program to institutions in the periphery.
- The program helps to stimulate connivance in institutions.

Administrative Staff and Principal.

- More time for the development of workshops with students, parents, and teachers. More frequency in the conciliation activities (at least once a month).
- Dialogue campaigns should be more frequent.
- The teachers in leadership positions should have more training.
- The focus group should be in both study groups and they should cover at least 10% of the students.
- It is necessary to train primary school children (mainly fourth and fifth grades) and some teachers who want to integrate [to the program].

Parents.

- Thanks because what I have learned allows me to work in a foundation for helping and guiding people.
- The program should not only be oriented to the students but also to the teachers because they must be people with qualities such as patience, love, authority and good leadership but because of their personal problems, daily stress and responsibilities they can hurt those who they are teaching.
- Directing care offices attended for and directed by parents as they are the root of many of the problems of their children.
- To turn the program in a transversal project to the educational and cultural activities.
- There is a lot of disbelief and distrust towards the project.
- Reconciliation is required as an alternative to conflict management, privileging dialogue, meeting and respect over violence, exclusion and marginalization, insults, and bullying.
- The consultants are very good but they work with few resources and dedicate very little time.
- Look for inter-institutional spaces to carry out campaigns at a global level that allow people to know the program.
- That the program can be extended to homeless population.

Graduates.

- It is important to improve communication among all the actors to develop more confidence in the program and to improve not only the relationships of others but their own.

Consultants.

- Greater training for consultants on issues that are relevant in the institutions.
- The intervention could be made more flexible in order to respond effectively to the needs of each institution without having to think about the time requirements of the activities proposed by each line of the program.
- The increasing projections in goals and indicators results in the weakening of the contents for the times that it requires to cover the phases of the plan. The administrative times are increasing and the fulfillment of the goals is intensified, which leads to a consultancy tending to respond to a protocol, procedures, and essentially administrative parameters. Some schools demand more content, quality, and accompaniment in the connivance processes than in formal procedures. The quality factor highly depends on the profile of the consultant and a logic of reducing and simplifying administrative procedures that improve the quality of the service.

Students did not have any additional comment.

In general terms the comments of the different actors were directed towards: increasing the frequency with which the workshops and activities of the program take place, widening the type of stakeholders who are trained by the program (not only by number but also by the actors that are being trained), recognizing that there are trusting issues regarding the program inside the institutions, and recognizing that although there is a high level of commitment from the Chamber of Commerce their personnel has to work with too limited resources and time. In the particular case of the consultants, administrative requirements have a large impact on the impact of the program as they are very time consuming.

9.1.2 Confidence Activity

One of the biggest constraints of conducting this type of workshops is power. Among the participants who were part of the workshop we could find two forms of power: role related power and age related power. The first one is related to the roles each of the individual plays within the school's community and within the program, whereas the latter refers to the power related to demographic characteristics between the participants, for example; age or level of studies related. For this reason, I decided to use an activity to boost confidence among the participants so it was easier to carry out the group activities which comprehended 80% of the activities of the workshop. The name of the activity I decided to use is "Crazy Trains" (its

methodology was explained in the evaluation planning chapter).

Although the results of the activity cannot be quantified, it was very good to appreciate how the participants of the workshop enjoyed doing it as it was something different to what they are used to. When asking the participant about the activity, I could say that the goals with which it was carried out were accomplished as they got to meet with a relaxed environment the people which they were going to conduct the other activities.

9.1.3 Drawing the community

Each group was asked to draw the way they felt the community was involved in the program and how they interact to achieve the desired goals. One of the most significant results about this, was the fact that several of the groups widened the boundaries that were initially placed around the actors considered as being part of the community. Originally I along with the director of the Program had considered the actors of the program as being the teachers, the principal and the administrative staff of the schools', the students, the former students of the program who have already graduated from the school, parents, and the consultants of the program (consultants work for the Chamber of Commerce and are the ones who train teachers and students in the Hermes Program). However, the participants of the workshop (which were the ones I originally considered) thought that other important actors of the program were the janitor and the maintenance team of the schools.

Below the reader can find pictures of some of the drawings done by each of the groups. I will explain them separately in more detail.

Figure 16. Group 1 Community Drawing.

Group 1: The human being is the central part of the school conflict. We wanted to portrait the diversity of roles in the group through the hands that appear in Figure 22, these roles are linked through the Hermes program. The faces correspond to the happiness of the children who have been able to solve their problems through the program and improve their school and community environment.

The support of the program does not only comes from inside the schools but also from the outside from the former students of the program who have already graduated from high school. The hands represent the support, creativity and work everyone does. Other students seek for the help of the conciliators to solve their problems.

The poster does not include a leader because somehow everyone has a leadership role in the conciliation tables, there is not a unique leader. The idea is that there is a healthy connivance among the different actors, this is illustrated through the happy faces in the poster.

One of the participants shared the experience with the program in the school he worked with, as follows:

The school, consisting only of elementary school, was founded in 2007. In 2008 and thanks to the law, the institution had to be extended to include middle school and high school. For this reason new students from sixth grade and up were assigned to the institution. However, the students who were assigned to the school were those who had a very poor performance in terms of coexistence/connivance/behavior in other schools. The program arrives in 2011 at the school and it represented a "salvation" to the school to be able to work with the problems of coexistence that were taking place. Today coexistence/connivance has improved by 80% comparing what it was before 2011 and what it is today. The institution has put a lot of effort to make the program part of the institution's coexistence/connivance manual, but this is something that has not yet been achieved.

Figure 17. Group 2 Community Drawing.

Group 2: In the center of the poster shown in Figure 23, is plotted the H that correspond to the Hermes program and around it are the different actors the group considers as fundamental part of the program; these are: students, the community surrounding the school (environment of the institution), the maintenance and cleaning staff (this includes the janitor and the cleaning services), teachers, secretaries, the administrative staff and principals, and the parents. The H is connected to each of the stakeholders through pathways with different shapes and distances that represent the ease with which each stakeholder can be reached. The depicted relationships are expressed as follows:

- The relationship with the directives is very important because they are the ones who decide whether or not to implement the program in the institution. If the chamber of commerce professionals do not generate appropriate relationships with these people, they will most likely not allow them to implement the program.
- The program is very important for the parents, they would like the program to work more extensively with them (the parents) because in this way it is possible to improve the relationship with their children. Sometimes there are difficulties with the parents because many of them are not linked to the program.

- One can appreciate a clock in the drawing corresponding to the teachers. This clock represents the time restrictions teachers have to get involved in the program. The pathway between the teachers and the program is built from the self, once a teacher joins the program, it will motivate others to join as well up to the moment they form a team. Another important element from the teachers' drawing are the text clouds which represent the different types of comments that come from them when hearing about the program, they could be positive, negative or they might not give an answer at all.
- The pathway to the maintenance and cleaning staff is filled with obstacles and it is not straight because it seeks to show that although the maintenance staff shares the physical spaces with the rest of the community their type of contract with the school and the dynamic of what they do is completely different. For this reason, their inclusion in the program is not that extensive.
- Students have the shortest and straightest path among all the stakeholders because they are the one that work directly with the program, in two specific ways: the student as conciliators, an activity that allows them to improve their interpersonal skills and to grow as human beings), and the student as conciliation clients. Although in this particular case they are not being certified, they learn to solve their problems in a different way from how they will usually do it.

The group also highlights the importance of passing the word about the program, as this is a fundamental way in which people decide to become part of it. In general terms students tend to be very aggressive and they get frustrated very easily. However, the students who are part of the program feel that they have been transformed. The group argues that this is an important reason to work with students since they are in elementary school until they graduate from high school.

Figure 184. Group 3 Community Drawing.

Group 3: Figure 24 shows the poster drawn by group 3. The main elements of the poster are a bridge and the people's faces. As peace makers, they do not only learn about conciliation but also to use dialogue as a bridge to achieve a long lasting peace. When people learn to use dialogue to solve problems, they not only become peace makers to solve problems but also to solve issues regarding their relationships at home. This means that they use dialogue wherever they pass by (symbolically represented by the bridge). The bridge also represents how the program brings together all these individuals with different backgrounds, features and cultural traits.

The Hermes program changes the culture not only in the school's community but also in the community in general which can also be expanded to the society and why not to the national sphere of our country, which is currently going through a peace agreement.

The image also shows the importance of using voice to voice to let everyone know about the program. Particularly, the group decided to use the case of one of the participant moms who started talked about the program to other people, introducing a single person first and then more and more until it reached a wider group of people. For this reason the poster shows the phrase "Fuerza Hermes", its translation is "Hermes Force" and it means that all the participants in some way become the army of the program because they work together to achieve its goals and they also attract more people through the positive impact it has had in their lives.

Figure 195. Group 4 Community Drawing.

Group 4: This group wanted to represent the program and its relationships with a tree (Figure 25). There is a small dot close to the smaller tree, this dot is a seed that represents an idea which one day arrived to the school(s) (the idea is the Hermes Program). The seed was planted by the consultants of the program but the roots of the program were given by the teachers. The reason why teachers are the roots is that they are the ones who provide the adequate spaces and times for the students to carry on with the activities of the program. The trunk of the tree represents the students, because they are the ones who keep the program alive, as they are at the center of it. The branches and the fruit of the tree represent the lives the program has changed. The fruit also represent the ideas left in the students' head every time they attend a conciliation session, as they are learning indirectly different ways of solving problems.

The program is a helping hand that helps solve problems without getting to an aggression. Previously, when a conflictive student left the school it was perceived that the problem had left as well. However, the student was carrying it as well. Nowadays that perception has changed, and problems are tried to be solved at the schools, by conciliations and seeking to transform the students.

Figure 206. Group 5 Community Drawing.

Group 5: This group also decided to make a representation by using a tree (Figure 26). The roots of the tree represent the values. Every human being is taught about values, but it is also well known that it is easy to be taught than to put in practice what is learned. Hermes is a project that helps and gives tools based in those values, such as: love, respect, tolerance, and patience which are fundamental values. If values are put into practice, the trunk of the tree will appear. The trunk represents the students, teachers, administrative staff and principal, the maintenance and cleaning staff, and parents of the schools, they are the people who strengthen the program. These actors do not only represent themselves but organizations as well (for example; the Chamber of Commerce, Universities, Schools etc.) and by strengthening themselves they will be able to strengthen the organizations as well.

The fruits are the results of the program in people's lives. They are primarily directed towards consolidating abilities in them such as: listening, be passionate, be more confident, patience, and others. The tree also has a watering can which represents bonding, building up, and

uplifting so that the program can be consolidated.

In the upper part of the poster there is a set of individuals with their hands tied, this is a representation of the diversity of roles that were represented in each of the groups and which constitute the variety of roles that can be found in the program as well. The land surrounding the program represents the institutional support; without it will be very difficult to conduct the proposed activities.

9.1.4 Small group activity

In this activity, the participants were asked to share changing experiences at an individual, institutional, family and community levels, related to the implementation of the program. The first part of the activity was done in the same small groups of the rest of the activities. The second part consisted of sharing the experiences which is why they voluntarily decided who wanted to share their own experiences.

The main highlights of the shared experiences were:

- The most significant experiences came from the parents, as they expressed situations were what they had learned from the program was useful to intervene in personal situation.

9.1.5 Six Thinking Hats

In the six thinking hats activity participants were asked to propose ideas about things that could be introduced or changed in the program. In the same groups they developed the previous activities, one participant would share an idea and the other participants of the group would adopt a particular hat and express their opinion according to it. This was done so that what only mattered were the individual's ideas and not their roles (this was related to handling power as well). Initially the idea was that all the members of the group would adopt every hat color in different rounds, however, due to time restrictions the activity could only be done once or twice in each of the groups.

In what follows I will present each idea with the corresponding arguments differentiated by color.

Idea 1

Green Hat (Corresponds to the proposed idea): Take the time the children spend in the program as social service.

Black Hat: Some teachers and students do not agree.

Yellow Hat: The children, parents, and the community are motivated to be part of the program.

Red Hat: It gives an opportunity to strengthen the interpersonal relationships, aside of that it motivates the students and makes them happy.

White Hat: It is good.

Blue Hat: It is a good proposal that many students like, it is easy to help teachers by validating their time just as it could be done with the students.

Idea 2

Green Hat (Corresponds to the proposed idea): Finding a moving space with banners (for example: the corner of the library), it can be built with the support of the school's administrative staff and parents to generate conciliation to train as conciliators.

Generate impact activities from the conciliatory parents to show in the community the impact of the program.

Black Hat: Spaces to conciliate are not available.

Yellow Hat: It is good because it creates conscience starting with the family.

Red Hat: It is sad because spaces are given and assigned but they are not taken advantage of, and also despair from the teachers in assisting the parents.

White Hat: I propose a pilot test and depending on whether it is feasible or not according to the information or result, the project could be implemented or if it is negative, mistakes can be corrected.

Blue Hat: Real bonds between adults (which are stronger) are the ones that guarantee a real commitment above duty.

Idea 3

Green Hat (Corresponds to the proposed idea): it would be good to have more fun activities, and more incentives for students and teachers as for example: fieldtrips, jean day, vests, and agendas.

Do sensitization events that empower the program in the institutions.

Black Hat: It is not a good idea because they are dangerous activities and there is no budget for transportation.

There must be available time to do the activities.

Yellow Hat: It is a very good idea as it is a motivation method to keep going with the program.

It is an excellent idea, it would be good to have theatre activities, magic, circus activities and others.

Red Hat: Children would be really, students would be very motivated as they feel important and recognized.

White Hat: There are approximately 40 students and 2 teachers per school. The transport is around 100.000 COP payed by the school, the park ticket would be around 25.000 payed by the Chamber of Commerce and 40.000 COP for each vest. The total budget would be around 1.600.000 COP.

Blue Hat: The different stimuli would be arranged according to the willingness of the Chamber of Commerce and the schools to strengthen the program.

The student's training and the empowerment means used in each institution should be dynamic, creative, and transformative, so that the school community will be able to see and appropriate with them.

Idea 4

Green Hat (Corresponds to the proposed idea): Modify the training system of the Chamber of Commerce towards using more innovative resources.

Doing parties, jean day, and other types of activities where alternative conflict resolution techniques are tackled.

Black Hat: The most recent events conducted by the Chamber of Commerce in the schools have kept the same development scheme.

There might exist conflicts with the administrative staff for the realization of the activities.

Yellow Hat: The topics are interesting and beneficial for the community.

The youngster need playful activities.

Red Hat: The events are very repetitive demotivating the participants to attend them. Sometimes the training gets boring.

We feel fun and happy in this type of activities.

White Hat: 4 events using the herringbone structure.

6 events using table rotation methodology.

1 specialized training.

Blue Hat: The different stimuli can be applied if there is a political willingness from the Chamber of Commerce and the schools to strengthen the Hermes Program.

The student's training and the empowerment in the project should be dynamic, creative, and transformative, so that the community so that the community will be able to see it and appropriate from it.

Idea 5

Green Hat (Corresponds to the proposed idea): Involve parents in the formative process: widen the levels of training, use of dynamics and children inclusion, working with external supporting networks.

Black Hat: Take into account the available time and spaces of the parents.

Hire workforce that can attend parents on weekends.

Yellow Hat: This motivates parents, voice to voice communication allows the transformation of people.

Red Hat: Frustration, demotivation in the training schedules, satisfaction for goals achievement.

White Hat: Low participation of parents in the Project [Hermes Program], seek strategies to keep the parents participation in the program.

The community close to the school is not directly linked with the Project.

Blue Hat: A coordinated work between parents, consultants, and teachers is important to keep the parents participation within the program.

Idea 6

Green Hat (Corresponds to the proposed idea): Workshops should be developed jointly with parents, students, and teachers.

Black Hat: There is no disposition from students or parents.

Yellow Hat: It is good because it is a way for actors to come close and there would be a unification of criteria.

Blue Hat: It is important for the unification of criteria so that it would be easier to achieve a solution for the conflict.

Idea 7

Green Hat (Corresponds to the proposed idea): Create a fight club to reinvigorate conflict resolution, a pillow fight would be a good idea as well.

Black Hat: Do not agree because it leads to not taking it serious and to aggression.

Yellow Hat: Great idea to release all the feelings and emotions, and in that same day do a vaccination against violence.

Red Hat: It allows the release of feelings and emotions.

Blue Hat: It is important to do it with well-established rules.

Idea 8

Green Hat (Corresponds to the proposed idea): Institutionalize a day for the Hermes Program.

Black Hat: No because in Colombia there are lots of activities that lead to nowhere.

Yellow Hat: It is positive because it can become an excuse for connivance.

Red Hat: Do not think it is right.

Blue Hat: It would be important if it has clear goals and objectives.

9.1.6 Overview of the results of the workshop

The participants were highly committed to conducting the activities of the workshop. The overall results have several aspects in common:

- Although the program places at its core the student's training, the stakeholders highlighted the importance of placing the parents, the teachers and the administrative staff and principals in the center of the program as well.
- The stakeholder's commitment was very variable, but in general terms the student's commitment was placed as being the highest among the considered stakeholders and the administrative staff and principal's one was perceived as being the lowest.
- Having the adequate spaces and times to conducting the activities of the program was perceived as a big obstacle for teachers and students.
- Administrative requirements were perceived as having a negative influence in the quality of the impact of the Hermes Program.
- Incentive packages and dynamism in the activities and workshops of the program were repetitively mentioned by several of the stakeholders as a way of motivating students and teachers.
- The success of the program is perceived as coming from the commitment and the work of all the members of the community, not only the consultants or the students.

9.2 Student Surveys Results

Taking into account the results of the workshop, two surveys were designed; a student and a consultant survey. The main reasons for this to take place, were the time and resource restrictions surrounding the evaluation. The student survey consisted of 12 multiple choice questions. However, 3 of these questions also had sub-questions making it a 34 question survey. The reader will find the questionnaire in the Appendix A of the document. As the questionnaires were written in Spanish, the reader will find the Spanish and the English version of each of the questions.

In total 1065 questionnaires were answered in 30 different schools. The students that participated in answering the questionnaire included a control group and four different classifications related to the time the students have been part of the program. This will be explained throughout the survey analysis. I will first present a descriptive analysis for each of the questions, this comprises a histogram, a frequency table, and descriptive statistics (where necessary). Taking into account these results, deeper and specific analyses such as hypothesis tests will be presented where necessary.

9.2.1 Grade

Figure 217. Grade Distribution.

Figure 27 presents the distribution of the students throughout school grades. As can be seen, the majority of the surveyed students belong to 8th and 9th grades with a 26.46% and a 28.81% respectively. This is mainly due to the fact that the program is primarily directed towards training students in these grades as they have an appropriate level of maturity (according to their age) and also because they will be able to complete all the proposed stages of the program before graduating from high school. However, a relatively large proportion of students can be found in other high school grades as well. The marked difference between the level of participation of students in 5th grade (0.66%) and high school students is due to the fact that although the program has a direct focus on students belonging to 8th and 9th grade, it is primarily oriented towards high school students and not elementary school⁶.

⁶ In the Colombian education system 5th grade is part of elementary school and from 6th grade students are considered as being in high school. Differentiating between middle school and high school is rare in public schools.

9.2.2 Age

Figure 228. Age Distribution.

According to Figure 28, age distribution seems to be following a normal distribution with a minimum value of 10 and a maximum of 21, where 23.56% of the students are 15 years old making it the most frequent value, and 0.09% of the students are 10, 20 or 21 years old. These three values can be considered atypical for several reasons. First, usually a high school student is as young as 12 years old, however in this case we are also taking into account 5th grade students which should be around 11 to 12 as well. On the other hand, it is expected that a high school student is not older than 18 years old. However, due to social and economic conditions, students might take longer to finish high school. Second, a possible reason for a student to be older than 18 is that he/she might have had a low academic performance, giving as a result failing and retaking a school year. Although this data is atypical, it will not be removed from the analysis.

9.2.3 Gender

Figure 239. Gender Distribution.

Figure 29 show the distribution of participants in terms of gender. As can be seen, 42.75% of the surveyed students are males and 57.25% are female.

9.2.4 Distribution of students through the program phases

Figure 24. Distribution by program phase.

In order to analyze the data, students were divided in 5 different possible phases. The first phase represents the control group, meaning that these students go to the same schools as the ones participating in the program, but they do not participate in it. The second phase corresponds to those students who are participating in the program but have been in it less than 6 months. The third phase corresponds to those students who have been in the program for more than 6 months but less than a year. The fourth phase corresponds to the students who have been in the program for more than a year but that have not been certified, and finally the fifth phase corresponds to the students that have already completed their participation and therefore have already been certified. Taking this into account, we can see that (Figure 30) more than half of the surveyed students (57.49%) have already been certified by the program. On the other hand, only 3.29% of the surveyed students are close to be certified as they have been part of the program for more than a year.

9.2.5 Household composition

To ease the data analysis and the compute of the information, the different types of household compositions were classified in a scale from 1 to 5. However, this scale does not follow a particular order of importance, meaning that having a value of 1 or 5 only means that the student has different household compositions without one being better than the other. The following are the household compositions represented by each value:

Question: With whom do you live with?

- 1- With my parents.
- 2- With my parents and siblings.
- 3- With my parents, siblings, and other members of the family (grandparents, uncles or aunts, cousins).
- 4- Only with my grandparents.
- 5- Other.

Figure 251. Household composition.

According to the results (Figure 31) most of the student's households are primarily formed by parents and siblings, with a 57.82% of the students choosing this option. However, 20.81% of the students also express that their household is composed not only by their direct family (parents and siblings) but by extended family such as grandparents, cousins and uncles or aunts as well. This means that there are many cases in which households are extended as they consist of more than parents and children. This may affect the family dynamics in which the student lives.

9.2.6 Quality of the family relationships

Figure 262. Perceived Quality in the students' family relationships.

The quality of family relationships is measured on a scale from 1 to 5 where 1 is taken as having a very poor quality in the family relationship and 5 means having a very good one. Figure 32 shows the frequency with which each option was chosen by the students. It can be seen that 51.37% of the students consider their family relationships as good, whereas 36.85% considers it as very good. On the other hand, less than 1% of the students perceive their family relationships as being very bad. In order to identify how the students throughout the different phases considered for this research perceived the quality of their relationships at home a differential analysis was conducted.

Figure 27. Perceived quality of the students' family relationships according to their distribution throughout the program phases.

Figure 33 shows the mean of the perceived quality relationships among the students belonging to each one of the program phases that were taken into account. As can be seen, the average among all the phase of the program is superior to 4. This means that in general terms, students perceived their family relationships as being good. However, it is worth noting that the students who are not part of the program are the ones who perceive their household relationships as having a poorer quality when compared with the students in the program. For this reason, a hypothesis test was conducted over the differences of the mean of each phase over their perception regarding the quality of their relationships at home. Table 1 shows the results of such analysis.

Ho: The mean of the perception of the quality of the family relationships is equal throughout all the phases.

Ha: At least one of the phases has a different mean for the perception of the quality of the family relationships.

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	6.11777744	4	1.52944436	3.08	0.0154
Within groups	521.6949	1052	0.4959077		
Total	527.812677	1056	0.49982261		

Table 1. F-test for testing differences in the perception of the program at home given the students' phases.

Bartlett's test for equal variances: $\chi^2(4) = 4.9928$ Prob> $\chi^2 = 0.288$

Given the results presented in table 1, the p-value for the F-test was 0.0154. This means that there is enough statistical evidence to conclude that there are no differences in the means of the quality relationships between the groups of students. With regards to the program, this would imply that there is not a statistical difference in the quality of the family relationships of the students no matter if they have been received some form of training by the Chamber of Commerce or not.

9.2.7 How do you think is the perception of the Hermes Program in your school?

Figure 28. Program Perception in the school.

The variable Perception in schools refers to how the members of the schools' community perceive the program. This variable can take values between 1 and 5 with 1 corresponding to the program being perceived as very bad and 5 for it to be seen as very good. An additional option corresponds to the students not knowing how the program is being perceived in the school's community. As Figure 34 shows, 35.38% consider the program perception as very good, whereas 6.90% do not know how the program is perceived within the school. In general terms, around 86% of the students consider that the program is positively perceived in their schools. However, the students' responses were further studied by putting into consideration the program phase they belonged to.

Figure 29. Program perception in the school differentiated by the phase to which the surveyed students belong to.

Figure 35 shows the mean regarding how the students in the different program phases consider the program is perceived in their school. As can be seen, the lowest average corresponds to the students who are not in the program with an average of 3.23, whereas the highest average is given by the students who have been in the program for more than a year. In order to explore if there is a statistical difference between the students' perceptions taking into account the phase they belong to, a hypothesis test was conducted.

Ho: The mean of the perception of the Hermes Program in the school is equal throughout all the phases.

Ha: the mean of the perception of the Hermes Program is different for at least one of the phases.

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	89.8600694	4	22.4650173	15.16	0
Within groups	1553.02502	1048	1.4818941		
Total	1642.88509	1052	1.56167784		

Table 2.F-test for testing mean differences between the students' perceptions about the perception of the program in the schools.

Bartlett's test for equal variances: $\chi^2(4) = 71.7984$

$\text{Prob} > \chi^2 = 0.000$

Table 2 shows that the f-value for the test is 0.000 as it is lower than the level of significance (0.05) it is possible to conclude that there is sufficient statistical evidence to say that there is a difference between the means of the different phases with regards to the program perception in the school. In order to find out which phases are the ones with different means, a Tukey test was conducted.

Perception at school	Contrast	Std. Err.	Tukey		Tukey	
			q	P>q	[95% Conf. Interval]	
Phase						
2 vs 1	0.904284	0.1645796	5.49	0.000	0.454565	1.354003
3 vs 1	0.7086183	0.1393274	5.09	0.000	0.3279018	1.089335
4 vs 1	0.9388704	0.2320073	4.05	0.001	0.3049031	1.572838
5 vs 1	0.8992376	0.1180207	7.62	0.000	0.5767422	1.221733
3 vs 2	-0.1956656	0.1533736	-1.28	0.706	-0.6147638	0.2234325
4 vs 2	0.0345865	0.2407046	0.14	1.000	-0.6231466	0.6923195
5 vs 2	-0.0050464	0.1343139	-0.04	1.000	-0.3720634	0.3619706
4 vs 3	0.2302521	0.2241972	1.03	0.843	-0.382374	0.8428782
5 vs 3	0.1906192	0.1018131	1.87	0.333	-0.0875883	0.4688268
5 vs 4	-0.0396329	0.2116156	-0.19	1.000	-0.6178793	0.5386136

Table 3. Tukey test for mean differences with regards to the program perception in the schools.

Tukey results in Table 3 show that the mean for group 1 differs from the mean of the other four groups. Group 1 corresponds to those students who do not belong to the program. This results are given the column correspond to the p-value. The first four numbers in this columns are lower than the significance lever (0.05) and is for this reason why the mean of such groups differs. As Figure 3 shows, the mean related to their opinion about the programs perception at the school is lower than the other four groups. This will only imply that under the view of those who do not belong to the program, its performance is worse than under the perception of those who are in the program no matter the stage they are in.

9.2.8 ¿How do you think people you live with perceive the program?

Figure 30. Program perception at home.

This question sought to know how the program was perceived in the students' homes. According to Figure 36, 46.05% of the students consider that the program is perceived as "good" at their homes, making it the most frequent answer among the different options. However it is also important to consider the fact that almost 25% of the students do not know how the program is perceived at their homes or think that their families are not aware of it. This statistic is one of concern as the Chamber of Commerce carries out socialization meetings for parents and teachers so that they get to know what is being done with the program. Further analysis was conducted regarding how the students in different stages of the program considered it was viewed at home.

Figure 31. Program perception at home considering the phases in which the students are.

Figure 37 shows the perception mean corresponding to each one of the phases in which the students could be classified. Those students who are not in the program are the ones who think that the program's perception in their homes is lower than in the rest of the groups. This could be due to the fact that when the students are not involved in the program, their parents are least likely to be in the program as well. For this reason, if parents get to have a deeper insight regarding what the program does and how it operates, their perception about it may be higher.

Given that some differences in perception of the program at home can be seen, a hypothesis test for assessing such differences was conducted.

The hypothesis upon which the analysis was conducted are:

Ho: The mean of the perception of the Hermes Program at home is equal throughout the phases.

Ha: The mean of the perception of the Hermes Program at home is different for at least one of the phases.

Taking this into account I wanted to test if there was a statistical difference between the different groups the students could be classified into in relation with the way in which the program was perceived at their homes. The used test was an F-test which is based on the comparison of variances.

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	38.2594821	4	9.56487052	7.48	0
Within groups	1346.54959	1053	1.27877454		
Total	1384.80907	1057	1.31013157		

Table 4. F-test for the differences in the mean of the perception of the Hermes Program at the students' homes.

Bartlett's test for equal variances: $\chi^2(4) = 21.9173$

Prob > $\chi^2 = 0.000$

Table 4 shows the results of the F-test, giving as a result an F-value of 0.000. Given that such value is lower than the significance value (0.05). We can conclude that there is enough statistical evidence to say that there is a mean difference among the students' groups regarding the perception of the program at home. In order to identify among which groups there were any differences, further analyses were conducted.

In order to test the hypotheses, a pairwise comparison of means was conducted following Tukey's method.

Perception at home	Contrast	Std. Err.	Tukey		Tukey	
			q	P>q	[95% Conf. Interval]	
Phase						
2 vs 1	-0.3937984	0.1528848	-2.58	0.076	-0.8115576	0.0239607
3 vs 1	-0.5694598	0.1291471	-4.41	0.000	-0.9223556	-0.2165641
4 vs 1	-0.565227	0.2155212	-2.62	0.067	-1.154141	0.0236867
5 vs 1	-0.5806837	0.1095871	-5.3	0.000	-0.8801315	-0.2812359
3 vs 2	-0.1756614	0.1422209	-1.24	0.731	-0.5642813	0.2129586
4 vs 2	-0.1714286	0.2236005	-0.77	0.940	-0.7824191	0.439562
5 vs 2	-0.1868852	0.1247283	-1.5	0.564	-0.5277063	0.1539358
4 vs 3	0.0042328	0.2080923	0.02	1.000	-0.5643814	0.572847
5 vs 3	-0.0112239	0.0941401	-0.12	1.000	-0.2684625	0.2460148
5 vs 4	-0.0154567	0.1965522	-0.08	1.000	-0.5525374	0.521624

Table 5. Tukey test for the perception of the Hermes Program at home for each one of the phases in which students could be located.

Table 5 shows the results of the Tukey test. As can be seen in the graph correspondent to the generated p-values one can identify two cases in which the perception of the program at home are different. The first case corresponds to the difference mean differences between groups 1 and 3 given that the p-value (0.000) is lower than the level of significance. Group 1 corresponds to those students who are not in the program and those who have been in it between 6 months and 1 year. The second case in which there are differences is between

groups 1 and 5 with a p-value of 0.000. Group 5 corresponds to those students who have already been certified by the program.

9.2.9 To which degree does the student agree with each of the statements presented in this question?

In this question students were asked about to which degree they agreed with a series of statements. The question included 7 statements related to the time, space and topics of the program, as well as the appropriate level of commitment of the school's staff (teachers and administrative personnel).

All the statements had 5 possible options going from “completely disagree” to “completely agree”, these options were computed by using a numeric scale from 1 to 5, each option corresponds to a number as follows:

- 1- Completely Disagree
- 2- Disagree
- 3- Undecided
- 4- Agree
- 5- Completely Agree

The adopted numeric scale is intended to show that choosing a higher number is more favorable than choosing a lower one.

In order to analyze the statements, they were grouped in two sets, one involving statements related to adequate spaces, time, schedules, and topics and a second set assessing the statements regarding the teachers and the administrative staff commitment, and the treatment the students felt they were receiving from the Chamber of Commerce consultants.

First group statements:

The school provides adequate spaces for the development of the program.

The school provides adequate times and schedules to conduct the activities related to the program.

The workshops conducted with the Chamber of Commerce consultants have an adequate frequency and duration.

The topics of the workshops are appropriate.

Second group statements:

The principal and the administrative staff of the school are highly committed with the development of the program.

Teachers are highly committed with the development of the program in their institution.

The Chamber of Commerce consultants treat students well and they show enough knowledge about the topics of the program.

Figure 32. Mean of the students' perception regarding different issues related to the operation of the Hermes Program differentiating by the phase of the program in which the students are.

Figure 38 shows the mean for each one of the first group statements differentiating the students' response according to the phase of the program they were at, at the moment they answered the survey. For all the assessed issues, the students who had already been certified by the program scored them more positively than the rest of the students. On the other hand, those students who are not part of the program are the ones who scored such issues with the lowest values. For the case of the adequate spaces given for conducting the activities of the program, the mean of all the groups ranged between 3.43 and 3.79. This would mean that in general terms students are not completely satisfied with the space given to them for conducting such activities. The average for this issue is the lowest for all the issues that were considered in the analysis of this question, meaning that improvements could be made in this regard. On the other hand, the means related to the frequency with which the activities of the program take place, range between 3.71 and 4.41 making it the best ranked issue among the ones taken into account in Figure 37. As students make progress with respect to their

participation in the Hermes program, their valuation with respect to this issue increases as well.

On the other hand, the following analysis included the second set of statements presented earlier in this question.

Figure 33. Mean about the perceived commitment of other stakeholders with the Hermes Program taking into account the program phase in which the students are located.

Figure 39 presents the mean for the questions related to the students' perception about the administrative staff commitment, the teachers' commitment, and how students feel they are treated by the Chamber of Commerce consultants. As it happened in Figure 39, students who are not part of the program are the ones who assess the issues with lower scores than those who are in the program. Out of the three issues considered in the figure, the one with the highest performance was the one related to the treatment given by the Chamber of Commerce consultants', such values range between 3.70 and 4.52. On the other hand, in all of the cases the average score given to the commitment of the teachers and the administrative staff of the school did not surpassed a value of 4. This means that in the students' perception, their commitment is between being acceptable and good.

In order to explore the possible differences between the perceived commitment of both groups, the following hypothesis test was conducted. However, in this case, such test is not differentiated by the phase in which the students are at but only takes into account their

general perception in both issues.

Ho: The mean of the teacher's commitment is equal to the mean of the administrative staff commitment.

Ha: The mean of the teacher's commitment differs from the mean of the administrative staff commitment.

Variable	Obs	Mean	Std. Error.	Std. Dev.	[95% Conf Interval]	
Administrative Staff Commitment	1054	3.721063	0.295631	0.9597762	3.663053	3.779072
Teacher's Commitment	1054	3.851044	0.264355	0.8582397	3.799171	3.902916
diff		-0.129981	0.300947	0.9770358	-0.1890334	-0.0709286

Table 6. T-test for differences in the mean of the administrative staff and the teachers' commitment.

Ho: mean (diff) = 0		t = -4.3191
Ha: mean (diff) < 0	Ha: mean (diff) ≠ 0	Ha: mean (diff) > 0
Pr (T < t) = 0.000	Pr (T > t) = 0.000	Pr (T > t) = 1

The obtained t value from the t-test has a value of -4.3191, the p-value for the two tailed mean difference equals 0.000. Given that such value is lower than the level of significance (0.05) there exists enough statistical evidence to say that the mean of the teacher's commitment and the mean of the administrative staff commitment are different. From this test, we can also conclude that the mean difference is below zero (mean (diff) < 0) given that its t-value takes a value of 0.0000. In terms of the analysis, this would mean that the students perceive the administrative staff commitment as being higher than the teachers' commitment.

In general terms, answers given to the 7 statements show that when evaluating elements managed by the school and by the Chamber of commerce regarding planning, design and preparation of material for implementing the program, those elements in the Chamber of Commerce hands have a better performance than those related to the internal operation of the schools. Particularly, those related to the commitment of the authority figures within the school.

9.2.10 Frequency with which students are victims and promoters of different types of conflict

Both of these questions were analyzed in a joint manner so that their results could be compared as they referred to the same behaviors but considering changes in the students' position, the first one taking them as victims and the second one taking them as victimizers. In order to give a temporal frame to the question, the student were asked about the frequency

of such behaviors in the month before the survey was applied. The frequency with which behaviors take place were measured by using a Likert Scale using 5 different measurements (Never, Rarely, Occasionally, Very frequently, and always). Each of the possible options corresponded to a numerical range from 1 to 5. In both cases the numerical correspondence is the same. This means that “Never” will be represented by number 1 whereas “Always” will be represented by number 5. The larger the number the more frequent the student is victim of these “attacks” or the more frequent it is the perpetrator of them, according to the particular case being studied.

Figure 34. Frequency with which students see themselves as being victims of certain behaviors.

Figure 35. Frequency with which students have certain behaviors towards other students.

Figure 40 shows the frequency with which students see themselves as being victims of certain behaviors, whereas Figure 41 shows the frequency with which they see themselves as victimizers of the same behaviors. The idea of considering both figures at the same time is related to having the possibility of comparing the results, given that students are placed in both positions. First, we must place attention in the situations in which the student is placed as a victim. As can be seen, for all the stated behaviors, more than 95% of the students consider that they have been victims on a regular basis. However, such frequency changes according to the severity of the behavior. For example the behavior they see themselves as being victims more frequently of threats and physical abuse as more than 80% of the students consider themselves as “always” being victims of such behaviors. Furthermore, the least frequent behavior is the one related to being insulted as 32.6% of the students see themselves as always being a victim of it. The differences between the students

On the other hand, the students recognize insults and mocking as being the conducts they infringe more often over other students. However, the frequency with which such behaviors take place is lower to those identified by them in a victim position. 25.4% of the students recognized occasionally insulting others whereas 23.3% of the students recognized occasionally mocking at other classmates. Mocking can be seen as the behavior that takes place more frequently as 6.7% of the students have admitted to do it very frequently or always. Further, the behavior that takes place with the least frequency is theft as 93% of the students have never steal something from their classmates and only 0.1% admits to always do it. This trend in the data can be explained taking into account the fact that jokes, pranks and mocking are usually considered as being less serious whereas theft and physical abuse are seen as severe and more easily recognizable.

As can be seen, there are differences in the frequency with which behaviors take place when the student is placed as a victim and as a victimizer. There are several considerations regarding this perception that need to be taken into account. First, the students see themselves as being victims of the most serious behaviors from their classmates, but when being asked about the frequency with which they attack others, the most frequent behaviors are the least severe ones. The degree of severity which could be attributed to a behavior is given by degree in which the well-being and integrity of the student could be affected as a result of the behavior. Second, there are no differences in the behavioral patterns identified by the different student groups. Third, it may be more difficult to identify “simple” behaviors such as mocking or insults, than identifying severe ones, as the former are not regularly taken as an

offence but more as a joke.

In order to understand if there are more differences between the students, further analysis was conducted by considering the student's gender.

Figure 36. Mean with which students recognized themselves as being victims of certain behaviors taking into account gender.

Figure 42 shows the mean of the frequency for each one of the stated behaviors by taking into account the students gender. A general appreciation of the graph will show that both groups follow the same trend regarding the behaviors they feel they are victims of more frequently. In both groups of students, physical abuse and threats are seen as the behaviors which take place the most. However, in the case of females, such behaviors take place more often as they have a higher mean of 4.83 for both cases, when compared to 4.79 for threats and 4.66 for physical abuse. Moreover, in both cases the behavior for with they feel victims less frequently is that of insults with a mean of 3.95 for males and 3.94 for females.

Figure 37. Mean with which students recognized victimizing others differentiated by gender.

Figure 43 shows the mean for each behavior when students are placed as victimizers. In this case, some results vary more than did when both genders were placed as victimizers. In general terms, the figure shows that males see identify themselves as victimizing others more frequently than women. This is shown as the means for all the behaviors is higher in the male group than in the female group. However, the trend over the behaviors is the same in both groups. In other words, although there are differences in the mean of the frequency with which both groups of students insult others, this is still one of the most frequent behaviors of both of them. The largest mean difference between both genders is in the frequency with which they recognize mocking to be taking place. In the case of men such behavior has a mean of 2.25 whereas in the case of women such value is around 1.98.

In order to identify if there is a real difference between both groups in the frequency of mocking at others, I conducted a hypothesis test for mean differences for mocking taking into account gender.

The following are the hypotheses that were taken into account:

Ho: the mean with which male students mock at others is equal to the mean with which female students mock at others.

Ha: the mean with which male students mock at others is different to the mean with which

female students mock at others.

The hypothesis test was carried out using a t-test as follows:

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Male	450	1.566667	0.0393274	0.8342608	1.489378	1.643955
Female	605	1.557025	0.0323865	0.7966015	1.493421	1.620629
combined	1055	1.561137	0.0250148	0.8125011	1.512053	1.61022
diff		0.0096419	0.0509463		-0.0903395	0.1096233

Table 7. T-test for the mean with which males and females mock at others.

Diff = mean(0) - mean(1) t = 0.1893

Ho: diff = 0 Satterthwaite's degrees of freedom = 942.324

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.5750 Pr(T > t) = 0.8499 Pr(T > t) = 0.4250

Table 7 shows statistics for both groups (male and female), giving as a result a t-value of 0.1893 and a statistical significance of 0.8499 for the t-test. Taking into account a level of significance of 0.05 and that $0.8499 > 0.05$, there is not enough evidence that suggests a difference in the means of males and females. In terms of the evaluation, there does not exist a significant evidence in the frequency with which mocking takes place between males and females.

9.2.11 Student's behavior when witnessing conflicting situations between other individuals

In this case students are asked about their reactions when witnessing them as a third party (not as a victim or as a victimizer). In this case, the student is not asked about the frequency with which these situations take place, but about which would be its preferred course of action to solve it. Each option is represented by a numerical value. This is done only to present results in a clearer way as each one of the possible courses of action is very extensive. Each number corresponds to an option as follows:

- 1- I do not do anything, it is not of my business.
- 2- I express my disagreement but I do not get involved in the conflict.
- 3- I provoke dialogue between the parts.
- 4- I confront them and I fight with them.
- 5- I help the one I like the most.
- 6- I tell what is happening to the teacher.

Figure 38. Students' response when witnessing inappropriate behaviors among other students.

Figure 44 show the frequency with which each course of action was chosen by the students. The preferred course of action varied according to the severity of the behavior that was being tackled. At first glance, it can be seen that in all the cases less than 6% of the students choose to fight or help those who they like the most among those who were involved in the conflict. For situations involving mocking and insults, the preferred course of action was expressing disagreement without intervening in the conflict with 35.5%, and 33.5%. Moreover, in the cases involving more serious behaviors such as threats, physical abuse, fights, and theft, students preferred to call the teacher to intervene or solve the situation. Only in situations involving rejection the preferred course of action was to provoke dialogue, as 29.2% of the students chose this option.

Provoking dialogue is an important issue as it one of the core aims of the Hermes Program. For this reason it will be further explored in what follows.

Figure 39. Percentage of students who used dialogue as a response to inappropriate behaviors differentiated by gender.

Figure 45 shows the use of dialogue in each behavior according to the gender of those surveyed. As can be seen, the use of dialogue is more likely to take place in situations which do not put in danger the students' well-being such as those involving mocking, insults, and rejection. In such situations women are more likely to use dialogue than men as can be seen in the percentage differences shown in the graph. For example, when witnessing situations involving rejection, 21.85% of women used dialogue as a mechanism of intervention compared to 19% of men. However, in more risky situations where dialogue is less frequently used, men are the ones who use it more often than women. For example, in the case of witnessing physical abuse between other students, 14.49% of men would use dialogue as an intervention tool whereas 11.56% of women would do the same. In both cases, the behavior with which both, men and women, would use dialogue the least are those involving theft.

Figure 40. Frequency with which students chose to use dialogue in each one of the considered behaviors and differentiated by the phase in which the student could be in the program.

Figure 46 comprises 5 different graphs showing the frequency with which students in different phases of the program chose to use dialogue throughout the different behaviors that were presented to them. As in the other analysis conducted with respect to dialogue, the students use dialogue the most in situations that would not mean any risk for them. However, there are changes in the frequency with which they would do so. Out of the 5 considered phases, 20% of the students who have been in the program between 6 months and 1 year would use dialogue to solve situations involving mocking whereas the students in the other 4 phases, including those who are not in the program would do so more frequently to solve situations involving rejection. It is of particular interest to see the results of the students who are not in the program. Such students are the ones with the highest percentages of the use of dialogue among situations involving mocking, insults, and rejection. It is interesting because the students in the other 4 groups have received some training in dialogue and conflict resolution techniques, but those who have not turn out to be the ones using such tools more frequently.

In order to find out if there exists a real difference in the use of dialogue between the students that belong to the different phases of the program, I carried out a hypothesis test. As there are more than 2 groups represented by the 5 phases in which students can be grouped, the test was carried out by using an F test.

Ho: the average use of dialogue is equal throughout the phases.

Ha: the average use of dialogue is different in at least one of the phases.

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	12.9757054	4	3.24392636	0.9	0.465
Within groups	5367.85975	1484	3.61715617		
Total	5380.83546	1488	3.61615286		

Table 8. F-test for the use of dialogue among the students in the different program phases.

Bartlett's test for equal variances: $\chi^2(4) = 2.4231$ Prob > $\chi^2 = 0.658$

Table 8 shows the result of the analysis of variance produced by the F-test. The table shows an F-statistic of 0.465, given that 0.0465 is larger than the level of significance (0.05), there does not exist sufficient statistical evidence to reject the null hypothesis. As a result, the mean regarding the use of dialogue for the students of the different phases of the program is the same.

9.2.12 Aspects the students like the most and the least about the Hermes Program

For questions 10 and 11, students were asked to choose at most 3 items from a list of 7 options (the list included the same items for both questions) they considered as being the aspects of the program they liked the least and the most.

Figure 417. Students' preferences over issues related to the program.

Figure 47 shows the frequency of the issues like the least and the most of the Hermes Program. The two elements the students like the most about the program are related to the workshops conducted with the consultants (21.4%) and the Chamber of Commerce personnel, and the support they receive from the consultants (18.3%). On the other hand, the two elements they like the least about the program are their classmates' commitment (27.6%) and the school's commitment (19.7%) with the development of the program. Such results are consistent with those presented in question 6 where the students were asked to assess different issues related to the implementation of the program in their schools. Although some aspects had a better performance than others, in general terms the students perceive that the school's personnel commitment (teachers and administrative staff) could be better, as well as improving the organization of the space and time given to them for conducting the activities of the program.

It is possible to identify a pattern in the elements students like the most and the least about the program. Those elements students like the most, are related to their personal being, in the sense that the workshops conducted with the consultants, having the consultants as a supporting figure, and perceiving an impact of the program in their personal lives are elements directed towards the improvement of the individual which are elements easy to personally identify and take advantage of, as it is not necessary to have a contribution or the commitment of others. In other words, these are elements the student can take advantage of because it is something that only requires its own personal commitment. On the contrary, the elements students like the least are entirely related to the commitment of others, primarily with the way the program is handled inside their school.

9.2.13 How do you think has been the implementation of the program in your school?

In order to measure the general perception of the implementation of the program, I decided to use a Likert scale with 7 levels with the following options: dreadful, very bad, bad, acceptable, good, very good and excellent⁷.

⁷ As the question was originally written in Spanish, this is an approximation of the words originally used, as in some cases there is no a literal translation of the word.

Figure 42.Student’s general perception about the Hermes Program.

Figure 43. Students’ general perception of the Hermes Program differentiated by the program phases in which they could be. Figure 44.Student’s general perception about the Hermes Program.

Figure 45. Students’ general perception of the Hermes Program differentiated by the program phases in which they could be.

Table 9. Tukey test for the program’s perception among the students in its different phases. Figure 46. Students’ general perception of the Hermes Program differentiated by the program phases in which they could be. Figure 47.Student’s general perception about the Hermes Program.

Figure 48. Students’ general perception of the Hermes Program differentiated by the program phases in which they could be. Figure 49.Student’s general perception about the Hermes Program.

implementation of the program. Out of these results, 38.84% of the students consider the implementation of the program as being “good” with around 87% of the surveyed perceiving the implementation of the program positively and 10% perceiving it as being “acceptable”.

Figure 50. Students' general perception of the Hermes Program differentiated by the program phases in which they could be.

Table 10. Tukey test for the program's perception among the students in its different phases. Figure 51. Students' general perception of the Hermes Program differentiated by the program phases in which they could be.

Table 11. Tukey test for the program's perception among the students in its different phases.

Table 12. Correlation between the students' survey variables. Table 13. Tukey test for the program's perception among the students in its different phases. Figure 52. Students' general perception of the Hermes Program differentiated by the program phases in which they could be.

Table 14. Tukey test for the program's perception among the students in its different phases. Figure 53. Students' general perception of the Hermes Program differentiated by the program phases in which they could be.

Considering the general perception of the students in more detail, Figure 49 presents the mean of the general perception for each one of the student groups of the survey. The figure shows that the group of students with a higher perception of the program are those who have been in it for more than a year but who have not been certified yet. On the other hand, those who perceive the program's implementation in a lesser way are those students who are not part of it giving it an average score of 5.10 which is equivalent to considering the program as acceptable.

Ho: The average general perception of the program is equal throughout the phases.

Ha: The average general perception of the program is different in at least one of the phases.

Taking into account such results, I proceeded to conduct a hypothesis test over the means of each one of the phases with regards to their general perception of the program.

Analysis of Variance						
Source		SS	df	MS	F	Prob > F
Between groups		35.4022967	4	8.85057417	10.22	0
Within groups		900.010539	1039	0.86622766		
Total		935.412835	1043	0.89684836		

Table 15. F-test for the program perception for the students in each one of the program phases.

Bartlett's test for equal variances: $\chi^2(4) = 8.5557$ Prob> $\chi^2 = 0.73$

Taking into account the results presented in Table 9, and that $0 < 0.05$ which is the level of significance considered for the analysis, there is sufficient evidence to say that there is a statistical significant difference in the general perception of the program between the 5 student phases.

Given these results, I proceeded to make a pairwise comparison between the means of each phase.

General Perception	Contrast	Std. Err.	Tukey		Tukey	
			q	P>q	[95% Conf. Interval]	
Phase						
2 vs 1	0.4598939	0.1258298	3.65	0.003	0.1160549	0.8037329
3 vs 1	0.5742686	0.10664	5.39	0.000	0.2828671	0.8656701
4 vs 1	0.7200443	0.1773818	4.06	0.001	0.2353355	1.204753
5 vs 1	0.5375497	0.0903386	5.95	0.000	0.2906928	0.7844065
3 vs 2	0.1143746	0.1173683	0.97	0.867	-0.2063429	0.4350922
4 vs 2	0.2601504	0.1840314	1.41	0.619	-0.2427288	0.7630296
5 vs 2	0.0776557	0.1027828	0.76	0.943	-0.2032058	0.3585173
4 vs 3	0.1457757	0.1714832	0.85	0.915	-0.3228147	0.6143662
5 vs 3	-0.0367189	0.0781234	-0.47	0.990	-0.2501967	0.1767589
5 vs 4	-0.1824946	0.1618502	-1.13	0.792	-0.6247621	0.2597728

Table 16. Tukey test for the program's perception among the students in its different phases.

Table 17. Correlation between the students' survey variables. Table 18. Tukey test for the program's perception among the students in its different phases.

Table 19. Correlation between the students' survey variables.

Figure 54. Consultant's gender distribution. Table 20. Correlation between the students' survey variables. Table 21. Tukey test for the program's perception among the students in its different phases.

Table 22. Correlation between the students' survey variables. Table 23. Tukey test for the program's perception among the students in its different phases.

Results shown in Table 16, show that there is a statistical significant difference in the General Perception of the program between group 2 and 1 ($0.003 < 0.05$), group 3 and 1 ($0.000 < 0.05$), group 4 and 1 ($0.001 < 0.05$), and group 5 and 1 ($0.000 < 0.05$). This would mean that there is not a significant different between the other phases. Bringing this results back to the analysis, this would mean that the perception of those who are not involved in the program is significantly different to that of those who participate in it without taking into account in which phase of the program they are in.

9.2.14 Correlation

Variables	School	Grade	Spaces	Workshop Topics	Victim of Insults	Victim of Mocking	Victim of Physical Abuse	Victim of Rejection	Victim of Threats	Observed Mocking	Observed Threats	Observed Physical Abuse	Observed fights	Threats to others
Age		0.728												
Time for Workshops			0.6123											
Treatment by Consultants				0.6164										
Victim of Mocking					0.5982									
Victim of Physical Abuse						0.5809								
Victim of Rejection					0.5447	0.5325	0.5445							
Victim of Threats							0.6434	0.5433						
Victim of Theft									0.5409					
Observed insults										0.6076				
Observed physical abuse											0.5999			
Observed fights											0.5019	0.5782		
Observed theft													0.541	
Consultants	0.9743													
Theft to others														0.5678

Table 24. Correlation between the students' survey variables.

The correlation analysis was conducted for all the variables considered in the student's survey. However, as there are too many variables and most of them are not highly correlated, in the following section I will only present those which have a correlation superior to 0.5 or inferior to -0.5.

Table 11 shows the highly significant correlations among the variables of the student's survey. The table shows the relationship between every pair of variables. However, I did not keep its symmetric character so that it was easier to show the correlation values. The variables consultant and school are highly correlated because the way in which the data was organized. This means that a higher number attributed to a school a higher number was attributed to the evaluator. However, this is not related to a level of importance but to an organizing tool. For the case of the age and grade variables, exists a positive relationship between both of them as it is expected to have older children in higher courses. On the other hand, the variable associated to the adequacy of the space given for conducting the workshops of the Chamber of Commerce is positively correlated with the time for conducting these workshops. In that sense, when adequate spaces for conducting such activities are given, an adequate amount of time for conducting them is given as well.

On the other hand, the variables related to the situations in which the students are victims are correlated as well. This would mean that when a student is victim of certain behavior, it is very likely to become a victim of another type of behavior as well. Although, this would depend on the type of correlations between the variables (see Table 11). For example, the variables related to be victim of threats is positively correlated being victim of physical

abuse. This would mean that if a student receives threats it is very likely that they will be followed by physical abuse from other students. Table 11 also shows correlation between the observed behaviors. In this case, some forms of behavior are correlated as well. However, their correlation does not refer to the frequency with which they occur but to the way in which they are tackled. If two variables are positively correlated, then they will most likely be tackled in the same way. That is, the student will use similar ways for solving both situations. Finally, theft and threats to others, that is being the victimizer, are positively correlated as well. This suggests that when a student threatens others, it will very likely steal from them as well.

This was a general analysis of the surveys answered by the students of 30 schools that are part of the Hermes Program in Bogota. The reader may find more statistical information regarding the Hermes Program in Appendix F and Appendix G. The following section will be focusing on the analysis of the surveys carried out with the consultants of the Chamber of Commerce.

9.3 Consultants survey results

In this section I will present the results of the surveys answered by the majority of the consultants of the Hermes Program. This survey was developed and answered by using google tools for surveys as time restrictions did not allow me to conduct them in a personal way. However, applying the survey by using a digital tool, and the fact that the number of consultants was not too large, gave me the possibility of not only using multiple choice questions but also open questions (something that could not be done with the students due to the large volume of data that was being collected). For this reason, this section will not only have statistical analyzes but also the most important elements of the answers given to the opened questions that were asked.

At the time the survey was conducted, the Chamber of Commerce had 48 consultants working with the Hermes program. However, 40 of them answered the survey anonymously. The survey included 20 questions, 15 of which were multiple choice questions and 5 which were open questions. Question include topics related to demographic, personal, program planning and program implementation aspects. Where possible, each question will be discussed separately and at the end I will give a brief summary to connect the given answers. The original consultants' questionnaire is in Appendix E.

9.3.1 Gender

Figure 60. Consultant's gender distribution.

As can be seen in Figure 50, women represent 80% of the consultants who answered the survey, whereas men only represented 20% of the sample. In terms of number of consultants, this corresponds to 32 women and 8 men among the survey respondents. A possible explanation for this could be related to the type of careers and professional curriculum needed for the task.

9.3.2

Age

The age among the consultants is very variable as the youngest of them is 24 and the eldest is 51 (Figure 51). However, 40% of the surveyed consultants were younger than 30 years old. The average age of the consultants is 32.675. This means that when compared with the student's age, consultants are young.

Figure 69. Consultant's age distribution.

9.3.3 Level of Education: Which is your maximum completed level of education?

Figure 78. Consultant's level of education.

In order to present the results of the maximum level of education of the consultants, some clarifications need to be made so that the reader easily understands the options. The options given were: high school, (having a high school diploma), technical (a semi-professional degree which is superior to a high school diploma and inferior to a bachelor degree), bachelor degree (having a professional degree), and graduate education (includes a magister degree, a PhD or specializations (this type of postgraduate titles are also given in the Colombian context)).

Given the results presented in Figure 52, it is possible to tell that all the consultants have achieved a formal education degree of at least a bachelor degree. A plus is that 40% of them has achieved some type of graduate education as well. This is useful for the Chamber of Commerce because almost half of the surveyed consultants have some form of graduate education, making them more prepared for doing their job.

9.3.4 Field of studies: What is your career field?

Consultants were asked about their field of studies. As it was very difficult to ask about a specific career, I decided to group them by categories. In total consultants had 10 options to choose from, however they only selected the ones listed below. The other options do not have a number as they were not chosen which is why they are not considered for the following analysis⁸.

- 1- Health and Medicine
- 2- Social Sciences
- 3- Sciences (Physics, Chemistry etc.)
- 4- Art and Humanities
- 5- Education
- 6- Other

⁸ The option "Other" refers to those fields not considered in the other 9 options presented to the consultants.

Figure 87. Distribution of the consultant's fields of studies.

As can be seen in Figure 53, the fields of studies of the consultants are diversified as they encompass 6 different fields. However, 52.5% of them have a professional background in the social sciences, whereas only 2.5% have a degree in sciences. A particular highlight of the results is the fact that 17.5% of the consultants have a profession related to health and medicine.

9.3.5 Time as consultants of the Hermes Program: How long have you been working as a consultant of the Hermes Program?

When asked about the time they had been working for the Hermes Program, the consultants could choose among 6 options. Each option corresponds to a number as follows:

- 1- Less than a year
- 2- Between 1 and 2 years
- 3- Between 2 and 3 years
- 4- Between 3 and 4 years
- 5- Between 4 and 5 years
- 6- More than 5 years

Figure 95. Time as consultants.

As shown in Figure 54, the period of time each consultant has been in the program is very variable as 22.5% of the consultants has been part of the program between 1 and 2 years and the same percentage of them has been in the program for a period between 4 and 5 years. 20% of them have been part of the program for more than 5 years. On the other hand, the lowest percentage of the surveyed consultants (10%) is new in the program as 4 of the consultants have been working with the program for less than a year. The main conclusion regarding these results is that although there is not a specific pattern in the time consultants have been linked to the program, a large percentage of them have been linked to it for several years. This is very useful for program planning, consolidation and know how.

9.3.6 Number of schools the consultant is working with

When being asked about the number of schools each of the consultants was working with, all of them answered the same number. The given answer was 8 schools. Although it is a single answer, the results are presented in Table 12.

Table 33. Number of schools per consultant

Figure 100. Consultant's perception about stakeholders' commitment with the Hermes Program.
Table 34. Number of schools per consultant

Figure 101. Consultant's perception about stakeholders' commitment with the Hermes Program.

9.3.7 What do you think is done in the schools to improve the communication skills among the students that are participating in

Figure 102. Consultant's average perceived commitment of the stakeholders of the Hermes Program.
Figure 103. Consultant's perception about stakeholders' commitment with the Hermes Program.
Table 35. Number of schools per consultant

Figure 104. Consultant's perception about stakeholders' commitment with the Hermes Program.
Table 36. Number of schools per consultant

the Hermes Program?

The answers given to this question were very diverse. However, in general terms they followed a pattern. The identified activities could be classified between those conducted by the school (independently from the Chamber of Commerce) and those that directly involved the Hermes Program schedule and personnel. However, the majority of the consultants only focused on the latter. When referring to those activities and spaces fomented by the schools, the consultants referred to the spaces given, for example, in parent meetings to socialize what has been done so far in the program and the spaces given to conduct the activities of the program. One of the consultants considers that schools do not give the adequate attention to developing this competence because they do not see it as being relevant.

On the other hand, when referring to the activities conducted by the program, the consultants referred primarily to workshops that involved the different actors of the school's community, by working not only in improving verbal communication skills but also by working on assertive communication, leadership, active listening, and abilities to organize information, to elaborate questions, empathy, empowerment, and emotion management. Although most of these activities take place in person, the consultants also try to promote the use of technological tools primarily in the communication with students and teachers.

The workshops and activities the program does to improve the student's abilities, are aimed at the students using them in conciliation activities, as being able to become a conciliatory figure is one of the final goals of the program. Fostering conciliation will ultimately create a dialogue culture in the school.

9.3.8 What do you think is the degree of commitment of the different members of the community with the program?

This question was directed towards assessing the perception of the consultants with regard to the commitment of different actors of the community with the Hermes program. In order to do so, a Likert scale with the following 5 levels of commitment was used; very low, low, medium, high, and very high. Each level was represented by a number between 1 and 5, where 1 represented “very low” and 5 represented “very high”.

Figure 105. Consultant’s perception about stakeholders’ commitment with the Hermes Program.

Figure 55 comparatively shows the results for the 5 stakeholder groups considered in the question. Consultants see themselves as being the most committed stakeholders with the program, as 70% of them perceived the consultants’ commitment as being “very high”. Putting aside the consultants perception over themselves, the most highly committed members of the school’s community with the program are the students, as the 30% of the consultants perceive the student’s commitment as “very high” and 60% perceived it as “high”. On the other hand, the least committed stakeholders with the program are the parents particularly because 90% of the consultants consider their commitment to be average or below than that.

Figure 114. Consultant's average perceived commitment of the stakeholders of the Hermes Program.

Figure 56 shows the mean of the perceived commitment for each group of stakeholders. This figure depicts the patterns previously described in which among the school's community students are the ones perceived as being the most committed and the parents are the least. However, this figure also shows how in the consultants' perception, the administrative staff commitment is low. The mean for the administrative staff commitment is 2.9, using a numeric scale this would mean that their commitment is perceived as being between "low" and "acceptable". This is particularly because the administrative staff of the schools are the ones who bring the program to the school and are the ones who give the teachers and the students' adequate times and spaces for conducting the activities related to the program.

9.3.9 What factor do you consider restrict the commitment of the different members of the school's community? What do you think can be done to increase their commitment?

Consultants recognize several elements that restrict the commitment of the different actors of the institutions. Most of the elements that were identified are transversal to several actors and are linked between them, for this reason I will present the factors without directing them to specific actors. According to the consultants, the factors that reduce the commitment of the actors of the community can be summed up as:

- Time and Spaces. The schools with which the program work are public schools. This means that they are directly linked with the Education Secretary, and in that sense they have to follow particular guidelines and work with different programs

at the same time. This generates a time restriction in terms of it being really difficult to dedicate the proper amount of time to conducting the activities of the program as the schools have to comply with other requirements as well and in most of the cases the schools have to rush the development of the activities of the program. However, the inadequate time (and space) dedication is not only related to the bureaucratic processes surrounding the schools but also to the lack of relevance the program is given, the lack of commitment of the actors of the institutions, and the disorganization inside the schools. This not only translates in the lack of time to carry out activities but also in the lack of time for teachers and administrative staff to get involved in the activities of the program.

- Relevance. This is maybe the most important factor that affects the commitment of the actors with the program. In general terms, connivance aspects of education are relegated from the student's education as education is primarily focused on academic aspects. In that sense, when the main focus of the program is relegated in the schools where it is being carried out, achieving an appropriate level of commitment from the different actors becomes very difficult. This also has an impact on the possibility of implementing the program transversally throughout the school and not only working with some teachers and students.
- Motivation. A general characteristic that covers all the actors is the lack of motivation. Most of the consultants argue that actors expect some sort of retribution from participating in the program, whether it is via a recognition, price or other things. However, currently none of them receives any of this, which is why they see it as doing things that take time from them but do not give them anything in return. Particularly in the case of the teachers, most of them see the participation in the program as a burden because it is extra work load that is added to what they usually do as teachers which is a lot of work already, for this reason not too many of them want to belong to the program. In the case of the parents, they like to play a "peripheral role" in their children education, mainly due to their daily life responsibilities and the family dynamics they have at home. In the case of the students, consultants attribute their lack of commitment not only to the need of having a retribution but also to difficulties paying attention to the activities as tablets, computers and cellphones are big distracting factors in the classroom, and if the activities do not seem appealing enough, the students won't commit with them either. In the case of the administrative staff and the principal of the schools,

their motivation is limited primarily due to the need of fulfilling too many requirements, and also to placing the connivance element at the bottom of the priorities to tackle at the school. Although the consultants do not have a limited motivation, they do get “frustrated” due to the difficulties they encounter in the relationships with the teachers and the administrative staff and the principal of the schools. One of the consultant states that the profile of the consultant is wrongly chosen and in that way their strengths are not properly exploited.

On the other hand, consultants also present ideas with which these limitations could be tackled. The general ideas are:

- The possibility of institutionalizing the program throughout the schools will increase the motivation of the actors of the school community to get involved in it. This means that conducting the activities of the program will be mandatory and in that way it will have adequate spaces and times, making it easier for teachers to work with it as activities will be more organized.
- Creating an incentive package. This could be a good idea not only to motivate the actors but also to achieve the annual projections of the activities and goals of the program in the institutions. In that way, not only the individuals will benefit but the institution as well. It is important to note that this package should be customized according to the interests and the necessities of each of the actors. For example in the case of the teachers, reducing the work load in other things they do will give them enough time to plan and design activities, as well to commit with the program. In the case of the teachers and students, it would be important for them to be recognize for the work they do, as well as other types of incentives.
- Plan meetings with teachers, administrative staff and the principal of the school before starting the academic year to discuss about the real possibility of conducting the activities of the program, achieving goals and to foster their commitment with the program.
- Strengthen the relationship between the consultant and the members of the school’s community. This is an important element because teachers play a pivotal role between the Hermes Program personnel and the students. If the relationships between these actors improves, it will be more likely that the results will improve as well.

- Listen the real needs of the communities. Sometimes the structure of the program idealizes the school leaving aside the real needs of the school community.
- Conduct activities to place the program in the scope of the institutions as well as emphasizing the importance of working in human relations so that teachers, administrative staff and the principal convince themselves of the importance of conducting this type of programs at an institutional level.

9.3.10 How much do you think the relationships among the following actors have improved?

Consultants were asked about the improvement of the relationships among different actors of the institutions they work with as a result of the implementation of the program. Improvement could be scored with 5 different options ranging between a very low and a very high improvement. In numerical terms, the lowest level of improvement was quantified by number 1 whereas the biggest improvement was scored with number 5. The results for each pair of relationships are presented in what follows.

Figure 123. Impact of the Hermes Program over the relationships between stakeholders.

Figure 57 shows the improvement in the relationships of different stakeholders involved in the Hermes Program. As can be seen the relationships that have improved the most are two of which directly involved consultants, as 47.5% of the consultants have identified a very high

improvement in their relationships with students whereas 25% have recognized a high improvement in their relationships with teachers. On the other hand, in the consultants' perception, the relationship that has improved the least is the one between parents and consultants with 40% of the consultants recognizing a low or very low improvement in their relationships. It is worth noting that even though the consultants do not always witness the dynamics of the relationships between the other stakeholders, their appreciation about them is relatively positive.

On the other hand, a criteria regarding the impact of the program in improving the relationships among the different stakeholders is based on two considerations. First, the scale used to measure the improvement of such relationships, and second the partial perception of the consultants regarding relationships they do not witness on a regular basis. When referring to the scale, the particular scale used in the survey was a positive one. This means that it only considered relationships to stay the same or to improve, without giving the consultants the possibility of considering the relationships between the stakeholders to have worsened. For this reason, even if the consultants perceive a relationship to have a little improvement from its previous state, this still represents an improvement. On the other hand, although the consultants are at the center of the operation of the program, they might not be able to witness the entire picture of the situation in the schools on their own as they do not spend a lot of time in each one of them. Instead, their perception is based on what they see and what they are told from other stakeholders. Taking into account these two elements and the results shown in Figure 56, one can say that relationships between the different stakeholders has improved due to the program. Even though they have done so in different degrees.

9.3.11 In your opinion which are the most important variables for a successful execution of the Hermes Program

In this question consultants were given 13 options from where they could choose 5 at most. As in the other questions, each option is represented by a number ranging between 1 and 13 corresponding to each option as follows:

- 1- Students and teachers available time.
- 2- Times and spaces given by the school for the development of the activities of the program.
- 3- Student's commitment.
- 4- Teacher's commitment.

- 5- Commitment and support of the administrative staff of the schools.
- 6- Teacher's skills.
- 7- Quality and content of the teacher's training.
- 8- Quality, clarity, dynamism and content of the material developed for students.
- 9- Socialization of activities and schedules.
- 10- Sensitization of the educative community.
- 11- Consultant's professional profile.
- 12- All the above.
- 13- Other.

The number is only used to ease the presentation of the information as the numerical scale and the order does not give a particular importance to any of the options.

Figure 132. Frequency with which consultants consider certain variables to be determinant for the success of the Hermes Program.

Figure 58 presents the results of the frequency with which each option was selected by the consultants. The first consideration that needs to be addressed is the total number of observations. As was previously stated, each consultant had the opportunity to choose more than one of the options, for this reason the total number of observations is 178 as some of them did select more and some did not. A second consideration, is the fact that 0.05% of the consultants selected the option 13 which refers to additional variables that were not listed as

options, when choosing this option each consultant was asked to write what they considered as being important as well, the answers given were:

- Give incentives to the teachers and students that are participating in the program
- Taking into account and boosting the professional profile of the consultant that is working with the program

Third, as can be seen in the results, the three most frequently chosen options were the teacher's commitment with the program, the appropriate time and spaces given by the schools for the development of the program, and the students' commitment. On the other hand, the least relevant variables for the success of the program under the view of the consultants are the teacher's skills, and the socialization of the activities and schedules.

Given these results, one can conclude that in the consultants' opinion, the success of the program is mostly related to the way in which it is managed and operationalized inside the school and not in the way the Chamber of Commerce designed it. This is an important issue particularly because the chosen success variables depend on the commitment of the school's community with the program and not with its formal structure.

9.3.12 Do you consider that the administrative requirements of the Chamber of Commerce (goals, indicators, paperwork, etc.) restrict the quality and the impact of the program?

The possible answers for this question were "Yes", "No", and "Not Knowing". However, question 13 did ask the consultants to justify their choices for question 12.

Figure 141. Impact of administrative requirements over the quality of the Hermes Program.

As can be seen in Figure 59, 47.5% of the consultants consider that administrative procedures of the Chamber of Commerce reduces the impact of the program in a practical way. Although this is the highest percentage among the answers, we cannot say the majority of the consultants have this opinion, as the percentage is not over 50%. On the other hand, 7.5% of the respondents did not have an opinion regarding the issue. In the next question justifications for these answers will be explored.

9.3.13 Please justify your answer to the previous question

Question 12 had 3 possible options; do not know, no and yes. For this reason, results for question 13 will be presented in terms of the arguments for each of the possible answers.

- Do not Know

Those who have given a “Do not Know” as an answer, considered that in some ways the administrative requirements do have a negative effect on the impact of the program and other they don’t. They understand that the administrative requirements are a way to control and keep a trace of what is being done in the program. However, they find difficult to do things such as filling forms and checking attendance all the time, as they believe this might have a negative effect on the credibility of the program. The consultants feel they are being pressured with the number of cases they need to attend, because sometimes this means that they should work more with children which are not the focus of the program only to achieve the required number.

- No

Consultants that gave “No” as an answer, consider that the administrative requirements are the base for an adequate development of the program as they work as a way of quantifying what is done in the fieldwork. These processes are at the center of the program’s management and in that way they see them as adequate. However, they do have two observations regarding the administrative requirements. First, although achieving and following up goals is fundamental, the consultants consider that the base lines of the program should take into consideration the specific dynamics and contexts of the institutions as this has an impact on how each school moves forward in the base lines. Second, some leaders need to improve the way they communicate the requirements as sometimes the way they explain them is confusing. On the other hand the consultants mention the Simasc Platform as being very

helpful.

- Yes

Consultants that see administrative requirements as having a negative effect on the impact of the program, have several considerations to support their view. First, they consider that the amount of the administrative requirements, paperwork, and signatures needed restrict the time they could be using when doing the activities of the program. Second, most of the time the consultants have to fill several reports that gather very similar information, they consider that reducing the amount of reports would increase their efficiency with the program. Third, something that has been highlighted in different schools is the need to avoid a constant consultant rotation as this would limit the continuity of the program and the attention could be focused on consolidating empowerment processes within the community. Fourth, consultants consider that emphasis should be placed in quality and not quantity as sometimes the quantity of conciliatory processes needed makes them include students for example from second and third grade which should not be participating in the program just to fill enough forms that evidence the achievement of the desired number of conciliations. Fifth, the Simasc platform has been prioritized relegating the importance of the consultants field work to a second place. Sixth, just as some of the consultants who chose the “No” option thought, the base lines do not take into account local contexts and dynamics of the institutions and for this reason sometimes it is very difficult to implement the program or to keep the pace with which the Chamber of Commerce wants it to be implemented. Seventh, the Chamber of Commerce places more attention to the administrative requirements than to the real work that is being done inside the schools, having achieved the goals in terms of numbers does not mean that the quality and desired depth is being achieved. Eighth, in the consultants own terms, it is very likely that at some point they will start doing things just to accomplish the administrative goals and not for the real objectives of the program, as they feel pressured to achieve them.

9.3.14 In your opinion, what is the importance of the Chamber of Commerce staff commitment in the success of the Hermes Program?

The importance of the commitment of the Chamber of Commerce staff was measured by using a scale from 1 to 5 where 1 represented very low and 5 represented very high.

Figure 150. Importance of the commitment of the Chamber of Commerce over the success of the Hermes Program.

Results shown in Figure 60, point out that the majority (80%) of the consultants consider the commitment of the Chamber of Commerce Staff as being very important, whereas only 2.5% of them consider it as not very important in the successful development of the program. Although the consultants perceive their commitment as being important in the success of the program, in question 11 the success of the program to central variables related to the schools and not to them.

9.3.15 Question 15

This question included a set of 4 questions primarily related to the role of the consultants in the program. Each question had four possible answers; yes, no, maybe and do not know. Each question will be analyzed separately.

Question 1: If the appropriate conditions were given, would you consider the possibility of working as a consultant of the program in 2017?⁹

⁹ The survey was applied in November of 2016, for this reason the consultants were asked about working with the program in 2017.

Figure 159. Likelihood of the consultants of working with the Hermes Program in 2017

As shown in the Figure 61, 97.5% of the surveyed consultants would keep working with the program for 2017. Only 1 of the surveyed consultants would have to think more about it. These answers are very promising in terms of achieving some continuity not only with what the Chamber of Commerce wants to do with the program but also with the ones working in it. These consultants, aside of knowing the program and being skilled to work with it, have established a relationship with the schools and their communities that new consultants would need to establish again.

Question 2: Do you see your work in the program as a life project in the long run?

Figure 168. Likelihood of the consultant for working with the Hermes Program in the long run.

When asked if they considered their work as consultants for the Hermes Program, 72.5% of the surveyed consultants considered it as a long run life project (see Figure 62). This is positive for the implementation of the program particularly for the know-how each one of the consultants has accumulated throughout the years, the relationships they have fostered with the schools, and the communication channels that have been established.

Question 3: Do you think that a frequent rotation of consultants affects the program negatively?

Figure 177. Perception over the impact of the consultant's rotation over the performance of the Hermes Program.

Figure 63 shows how the majority of the consultants (82.5%), consider that a high rotation of consultants has a negative impact in the implementation of the program. The answer to this question is congruent with those given to the previous two questions. It is very likely that consultants take this into account when considering working for the program for a long period of time.

Question 4: Do you think that the program would have a larger impact if it had a larger time dedication per school and a smaller number of schools?

Figure 186. Consultant's perception over the possibility of working more time with a narrower number of schools.

In this question the consultants were asked if they thought the program would have more impact if instead of focusing on including more schools they focused on working more hours with a reduced number of schools (without expanding the number of schools). According to the results (Figure 64), 42.5% of the consultants saw this as a good idea. This question was asked in terms of achieving more results when focusing in less schools at the same time. However, as the results show, not all the consultants see this as a good idea.

Major highlights of the results of these 4 questions can be summed up as follows:

- The majority of the consultants see their work as consultants of the Hermes program as something they would do in the long run.
- The majority of them consider that a high rotation of the consultants has a negative impact on the performance of the program.
- Almost half of the consultants think that the program would have more impact if it was focused in a narrower number of schools, instead it focused on working more time with the chosen schools.

In general terms, this means that there is certain consistency at least in terms of how the consultants perceive their work in the program as they do not only think in their personal convenience (in terms of the economical retribution of getting the job done) but also in

considering that a sustained work throughout the years has an impact as well. However, there is not a clear consensus in how this would be best done (in terms of focusing on less schools or not).

9.3.16 Which activities do you think would motivate parents to get more actively involved in the program?

In order to facilitate computing the results, just as in previous questions, options are represented by numbers between 1 and 9. However, in this case, aside of representing a particular option numbers do not have any importance connotation. Each number represents options as follows:

- 1- Workshops with leadership and entrepreneurship topics for parents.
- 2- Activities of family living and communication for children and parents.
- 3- Sending periodic reports to parents regarding the performance of their children in the program.
- 4- Public recognition of the children's work in the program.
- 5- Developing activities with members of the community surrounding the school.
- 6- Working jointly with other schools within the Hermes Program.
- 7- None of the above.
- 8- All of the above.
- 9- Other.

Figure 195. Frequency with which possible ideas for fostering parental motivation are chosen.

Before getting into detail about the frequency with which each of the options was chosen, it is important to mention that one of the consultants chose option 9 which refers to “Other” as a different activity to motivate parents to become more active in the development of the program. The consultant who chose this option proposed the following:

- This statement is directed towards making it mandatory for parents and schools to be involved in topics and activities related with connivance and to make alliances to spreading the word regarding what the program and their children do to improve their schools and the community.

Moving on to the general results, consultants had the chance of choosing more than one option they consider would boost the parent’s commitment with the program. For this reason the number of observations is 67, as some of them did choose more than one option. The most popular option among consultants was options 2 (32.84%) (See Figure 65). This option corresponded to developing activities with topics related to family living, coexistence and communication for parents and children.

As can be seen the most popular options to improve the commitment of the parents are not directed particularly to involving them but fundamentally to focus more on the children. The first activity is directed towards working on the relationships between parents and kids whereas the second activity is focused on giving recognition to the child. It seems that under the view of the consultants, when the student feels that it is at the center of the activity/program, this feeling will naturally transfer to the parent and in this way it might be possible to involve them, in other words, this will be a collateral result of an activity originally intended to something else.

9.3.17 Are you satisfied with the current structure of the program?

This question had 4 possible answers; “Yes”, “No”, “Maybe”, and “Do not Know”. Question 18 asks the consultants to justify their preferences for question 17.

Figure 204. Consultants' satisfaction with the current structure of the Hermes Program.

According to the results, 60% of the consultants are satisfied with the current structure of the program, whereas 22.5% were undecided, 15% were not satisfied with it and 2.5% did not know when being asked (Figure 66). Although more than half of the consultants are satisfied with the current structure of the program, the general satisfaction is considerably low taking into account that 25% of the consultants do not have a defined opinion about the structure of the program.

9.3.18 Question 18

Question 18 asked consultants to justify their response to question 17. The justification for choosing each of the answers will be explained in what follows.

- Yes

The consultants who believe that the current structure of the program is adequate, argue that the general structure of the program is very cohesive and organized which is why there is coherence in what the program is intended to do and the way it does it. Some of the consultants believe that the internal changes that have recently been taking place within the program (changing from zones to cores) are very helpful to improve the communication among the consultants and to exploit their personal and professional experiences. A couple of consultants also express the necessity of involving the leaders in the sensitization processes of teachers and administrative staff of the schools, as they have more experience and expertise in for example normative subjects, that could be very helpful when approaching the necessities of the actors in their local contexts.

- No

The consultants including those who are not pleased with the current structure of the

program, express concerns primarily with the communication among consultants, and the role of the leaders of the program. First, in their opinion the past structure of the program (the one that was used until 2015) eased the flow of the information between the consultants, contrary to what is currently happening in the program where sometimes not all the consultants are aware of information that concerns everyone (one of the consultants put as example the Forum that took place in 2016). Second, only some of the leaders give an adequate support to the counseling the consultants give to the schools and to the meetings they have with the teachers and the administrative staff of the schools, which is why they believe this type of support needs to be consolidated. This concern is the same one as the one expressed by those consultants who liked the current structure of the program. On the other hand, the consultants also consider that the program has too many administrative requirements which take a lot of time to fill. They also have to comply with 5 to 6 different leaders, and they feel this is very demanding. Finally, they feel that the work of the consultant is not being given the adequate importance by the Chamber of Commerce as it deserves.

- Maybe

This group of consultants express mixed reviews about the structure of the program. They recognize that the construction of the program is a team work that has been taking place for many years with the help of many individuals, among which are the consultants, and it is an effort that requires a constant evolution and adaptation. However, they consider that the large quantity of activities of the current structure of the program makes it difficult to achieve the proposed goals, in their own terms, sometimes quantity comes first than quality. On the other hand, they also believe that the program is structured to the point where they are not allowed to determine what topics should be worked with according to the real need of the tutors (teachers that are being formed) and schools.

- Do not know

The person who chose this option, did not understand the question which is why the option was chosen.

9.3.19 How do you perceive the general quality of the Hermes Program when comparing it with other programs focused on conflict resolution techniques?

The possible answers to this question were between 1 and 5 where 1 corresponded to the worst score and 5 to the best one. This question addressed the perception of the consultants regarding the general quality of the program when comparing it to other programs with the

same scope (Figure 67).

Figure 212. Consultants' general perception about the program.

In general terms, the majority of the consultants (55%) scored the program quality as very high as they gave it a score of 5, 35% of the consultants gave it a score of 4 and 10% of them scored it with 3 (Figure 67).

Variable	Number of Observation	Mean	Standard Deviation	Min	Max
Quality Perception	40	4.45	0.6774765	3	5

Table 42. Hermes Program quality perception.

As shown in Table 13, the average score of the program is 4.45 which corresponds to giving it a high score between what can be considered as high and very high in terms of quality. On the other hand, the standard deviation has a value of 0.6774 placing the values around the average with a single standard deviation between 3.78 and 5.

9.3.20 Additional comments

This was not a question itself but a space that allowed consultants to write whatever they felt important about the program and that was not asked in the previous question. Some of the comments are positive whereas others are directed towards aspects that can be improved. It is very difficult to condense the opinion of the consultants in shorter phrases as they all had something different to say, for this reason I will present the translation as well as the original comment of those consultants who wanted to say something (not all of them commented on

this question).

- The theme of connivance is very broad, however, the program has been directed towards building warm and harmonic scenarios and provide tools to its actors in: emotional intelligence, basic skills, and currently it is been enriched with ATENEA, which provides students with social entrepreneurship, which is a “hit” in innovation. The program is constantly renewed, adjusted and consultants are permanently trained to give the highest professional level of the market.
- I think it is important to rethink some phases and workshops that do not demonstrate the quality or impact that the program should have. Promote more activities that enhance the personal side.
- Taking into account Question 19, it is necessary to innovate in new methodologies, although the program is well positioned it does not mean that we have to neglect what we have, that is why it is important to rethink the program to keep being in a good place.
- I think the program needs to renew its methodology and follow up on the impact it generates. It must provide the time for students to really take a turn, make a change and become real conflict managers.
- Working as a consultant of the program makes me a peace ambassador, for this reason I am proud to be a part of the program and the results that the young people and teachers I work with have achieved.
- The program is widely accepted in each of the schools.
- It is a program that not only contributes to the school’s environment but also to the life project of many students and their family groups which is an achievement that the schools recognizes. However many of the schools and teachers would like to have more recognition through advertising elements such as pens and vests, that would let everyone know that they come from an institution that is working with the Chamber of Commerce. As they perceive that other programs with shorter duration do give more material to their participants.
- In the different schools in which I have worked throughout my tenure in the Hermes Program (ten years), this program has been chosen as a serious program for living together in a stable and formative way, with specific principles and values. It also enables the development of social skills in the main actors in the institutions, the youngsters linked to the process. I have no point of comparison

with another school program because I do not know them. The Hermes Program of the Chamber of Commerce of Bogotá is unique in its methodology, stability and commitment.

- It is motivating and satisfying to be a consultant of a comprehensive program that seeks to improve convivencia in schools. It's a school that teaches you to be a leader, tolerant of frustration, purposeful, to strengthen communicative and emotional skills.
- It is a very good program, has a short-term, long-term objective, if the institutions take it as it should be, it would be a program with a greater social importance, as well as a greater breadth of resources at the Chamber of Commerce level.
- The work of the program in the schools is very valuable, short course training in conflict resolution and related subjects such as legal subjects to strengthen the advice to the committees of convivencia is very important.
- The program is important for the institutions therefore I think it is important that there is also training for the consultants.
- In some cases the leaders do not know the institutional dynamics and the profile of the new generations, this is evidenced in the requirements and the perspective of the field work.
- I think that the Hermes program manages to pick up an area of educational communities that is often in the background, which is convivencia, which is perhaps more important than any other knowledge that can be acquired in a school. In that sense, being able to involve people with a different vision of education in institutions and focus on working to improve student education has no equal (at least in what I have learned) in schools in the city.
- The consistency with which the process has been done is good as it has kept the credibility in the program.
- I would like to suggest that we take into account a little more the fact of being able to provide some kind of incentive by the program at the end of the year to the teachers, students and administrative staff involved in the program, since there are schools that give the maximum to achieve a great level of empowerment of the program in the institution. On the other hand I also want to congratulate all the professionals in charge of the program because it is clear that it gives good results and generates great impact throughout the educational community.

- It is a process that favors connivance and the students who compromise the most with the conciliation activities are the ones who are in elementary school. It would be appropriate to be able to generate seeds of coexistence leaders in these grades that begin to strengthen skills for when they are in the adequate grades for training as promoter
- The Hermes Program is a tool that greatly contributes to improving the coexistence/connivance of the educational community.

As can be seen, the additional comments of the consultants are directed to a wide arrange of aspects. However, when reading them, several conclusions can addressed:

- Although the consultants consider that changes and improvements need to be made, they also recognize the importance and the impact of the program in the institutions they are working with.
- As in the previous questions, consultants highlight the importance of incentives that motivate and recognize the active participation of the actors of the program.
- In general terms, consultants want the leaders of the program (who are in higher level of the administrative scale of the program) to get more involved in the field work as they have knowledge that the consultants do not have.
- Some of the consultants consider that the previous methodology of the program was better than the one they are currently using.
- The necessities and interests of the school's community need to be taken into account so that the way the program is being implemented can be customized and in that way maximize its impact on the practical level.
- Consultants believe that their personal and professional skills could be better exploited with the activities of the program. They also feel that their work could be more recognized than it currently is.

9.3.21 Consultants Survey Correlation analysis

The complete correlation analysis for the consultant's survey analysis is in Appendix G. As for the student's surveys, only those values higher than 0.5 or lower than -0.5 are considered as having high correlation. These will be the values to be analyzed in this section.

Correlation	Administrative Staff Commitment	Teacher's Commitment	Student's Commitment	Student/Student Relationship	Student/Adm. Staff Relationship	Student/Consultant Relationship	Relevance of the Chamber of Commerce Commitment
Student's Commitment		0.5063					
Consultant's Commitment			0.6046				
Student/Teacher Relationship				0.5296			
Teachers/Administrative Staff Relationship	0.56				0.5903		
Teachers/Consultants Relationship						0.5787	
Parents/Consultants Relationship					0.6676		
Student's Satisfaction							0.5148

Table 43. Consultant's survey correlation analysis.

Table 44. Consultant's survey correlation analysis.

Table 45. Consultant's survey correlation analysis.

Table 46. Consultant's survey correlation analysis.

According to the results shown in Table 14, every pair of correlated variables has a positive correlation between them. All the variables have in common their relation with the perception of the consultants regarding the commitment and the relationships between stakeholders. According to the results, the student's commitment is a key element in the perceived commitment of other stakeholders such as teachers and consultants. This means that when the commitment of one of these stakeholders is perceived as being high, the other stakeholder's commitment is perceived as being high as well. On the other hand, there is a complex dynamic in the relationships between the stakeholders of the program. For example, the relationship among students is positively and highly correlated with the relationship between the students and the teachers. As these variables make reference to the improvement of the relationships, perceiving an improvement in the relationship in one of these pairs will directly

lead to an improvement in the relationship of the others.

9.4 Findings

There are several conclusions that can be drawn not only taking into account the consultants survey but the initial workshop as well.

- A general appreciation among the consultants is that they do not feel their work is being recognized as it should be. This is exemplified by the fact that in the results of the workshop, they were barely mentioned by the other stakeholders as a key element for the development of the program.
- The need to create more dynamic environments, activities, and workshops for the development of the program was mentioned in all the conducted activities.
- The teachers expressed in the workshop the scarce availability of time they had to conduct the activities of the program. The consultants in their survey realized that this was a major problem for the impact of the program, the idea that was being considered was allowing the teachers to work in the activities of the program while reducing their work load in other activities.
- In the workshop different stakeholders perceived that the administrative staff and the principal's commitment was very low. In the survey directed to both, students and consultants, the administrative staff commitment perception was among the lowest, it was the lowest among the students' perception and the second low among the consultants perception.
- According to the correlation analysis from the consultants' surveys, in order to boost the commitment of all the members of the educational community, one should try to boost first the commitment of the administrative staff and the principal. The way the variables are correlated show that their commitment is the source of the other stakeholder's commitment as well.

- Student's and consultant's commitment, and the available time and spaces to conduct the activities of the Chamber of Commerce, are the variables perceived as having a larger impact on the success of the program.
- Consultants feel that the new methodology of the Hermes Program restricts more their work with the program, because it has too many administrative requirements that need to be filled, this reduces the available time to conduct the activities with the students.
- Although the workshop and the consultant's survey show a large impact of the program in the students' behavior, this impact does not always translate to how students react in real life situations involving conflict among other students. According to the students' survey results, when witnessing different conflictual situations students tend to express their disagreement without intervening in the conflict or they prefer to call a teacher to intervene and solve it. Students prefer to express their disagreement without intervening when witnessing mocking, insults or rejection. Whereas, they prefer to call the teacher in situations involving threats, physical aggressions, fights, and theft. These situations are characterized by having higher potential levels of violence involved and are those in which the student feels that its integrity (most likely physical) could be negatively affected as a result of its intervention in such conflict. However, this does not mean that the program has not had any impact in the stakeholders, particularly over the students, as they have expressed the opposed throughout the evaluation activities I have carried out with them. Instead, this shows that the students feel more secure and confident intervening in controlled situations such as conciliation tables. This also shows that although the students may acquire abilities such as the ones promoted by the Hermes Program, there are other personality and contextual traits that would foster or hinder their intervention for solving conflicts. These traits go beyond what the program could offer, particularly because they first need to be understood, so that mechanisms to tackle them get to be developed. Thus, the program could be seen as adequate in terms of appropriation and internalization of concepts. But there are difficulties bridging the theoretical component with the practical one in an environment in which the student is not in control of the situation.
- Regarding the frequency with which students are victims and victimizers of certain behaviors, students recognize themselves as being victims more often than being victimizers.

- The different stakeholders have a very good overall perception of the program. However, they propose changes that could take place in its structure and dynamics.
- Administrative Requirements put so much pressure on the consultants that sometimes they fall into inappropriate conducts to fulfil the requirements. Particularly, when they are asked to do a specific number of conciliatory processes and they do not have enough students to cover the number, they ask elementary school children to fill the forms and sign to show the evidence to the Chamber of Commerce of the completed activities. This is not proper as the program is supposedly to be oriented towards working with high school children and not elementary school ones. However, they have a mixed arrange of students and teachers from both, elementary and high school.
- Consultants believe that the leaders of the program could have a more active role with the community. Leaders have more preparation that consultants in some topics and their help would be very meaningful in solving the communities' concerns and questions about the implementation of the program, as something the consultants feel unprepared to answer this type of questions.

9.5 Conclusions about the Hermes Program Results

Taking into account the evaluation findings of both, students and consultants, as well as the results of the workshop conducted with other stakeholders, one can conclude that the impact of the Hermes Program needs to be addressed at an individual, relational, and operational level and taking into account broader elements than those related to the use of dialogue (which was the initial purpose with which the evaluation was set). At an individual level, students and parents consider that the program has had a significant impact on them because it has improved their lives, particularly given the skills they have learned such as leadership, communication skills, empowerment, team work, and problem solving skills and that they did not have before getting involved with the program. At the same time, the knowledge acquired by the students has had an impact on their ability of recognizing when they are being victims or victimizers following inadequate behaviors, and they state that the program has “changed them”, however they do not state in which was this has taken place. Students also recognize the importance of the program at a personal level in relation with the opportunities it gives them of seeking future education opportunities as well as in developing their own entrepreneur ideas. However, they would like to have more recognition from their participation in it.

On the other hand, at a relational level, there are two considerations. The first one is related to the use of dialogue among the students. Although dialogue is an important element to address how the program has had an impact within the school community, when addressing how frequently students use dialogue as a means for intervening in other people's conflicts, it is not the most popular option among them. On the contrary, there are situation in which students who do not belong to any of the Hermes Program stages tend to use more dialogue than those who are involved in it. Given the results of the evaluation, it seems like the students are interested in learning the skills promoted by the program, however there are difficulties in putting the learned skills into practice. This is related to two key issues: personal well-being, and the severity of the witnessed behaviors. When students witness behaviors such as mocking and rejection between their classmates, it is more likely that they will directly intervene, whether it is by using dialogue or not. However, in other types of situations that could be understood as more severe, such as witnessing fights, physical abuse or theft, students are more likely to call a teacher or to not intervene at all. Both situations are related to putting the personal well-being first, which is in turn related to likelihood with which someone may get harmed as a result of their intervention. Students perceive that by getting involved in more severe situations may bring as a result an increase in the likelihood with which they may get harmed.

A second element related to the relational character of the Hermes Program is associated to the impact of the program over the relationships among the different stakeholder groups, as well as the commitment of the stakeholders of the community with the program's development. With regards to the impact of the program over the stakeholders' relationships, both consultants and students consider that the program has led to an improvement of the relationships, particularly those involving consultants, students, and teachers. However, the extent with which each one of the relationships have been improved varies widely. When consultants and students were asked about how they perceived the commitment of the community members with the operation of the program, they perceived it as being lower than their own, in fact within the school's community, consultants and students perceived the students commitment as being the highest among the community stakeholders. On the other hand, when other community stakeholders were asked about the commitment of the different stakeholder groups with the program, the administrative staff viewed the students' commitment as being among the lower and perceived their own as being high. A general perception over this issue is that the administrative staff of the schools and the students have

an opposed view over their commitment with the development of the program. This is directly related to the quality of the relationship between both stakeholder groups and how the activities of the program take place within the school. However, it is important to highlight that there is a sense of community not only between students and administrative staff but also involving other members of the community as well. This is important because, although administrative staff and students see themselves as opposed, they see themselves as being part of a larger system with parents, teachers, and consultants. This leads us to operational issues of the Hermes Program and its evaluation.

Regarding the operational character of the Hermes Program, there are major differences between the perception of the structure of the program and how it is implemented within the schools. Concerning the general structure of the program, students and consultants have a positive opinion regarding the current structure of the program. However, changes regarding the topics the program tackles, making activities more dynamic so that they may be more appealing for students, and comments about the consultants work load and “bureaucratic” requirement that need to be filled were made. The role of the consultants within the operation of the program is highlighted by all the members of the schools’ community not only in the surveys but in the initial workshop as well. The closeness in the relationship between the students and the consultants is fundamental for the success of the program, particularly because students see them as a guiding figure for their personal development as well. On the other hand, when being asked about their long term working plans, working for the Hermes Program in the long terms was a feasible possibility for almost all the consultants. This is very important not only for giving continuity to the program’s implementation but also for strengthening the relationships between the Chamber of Commerce and the schools by means of the work done by the individuals in the role of consultants, which are a bridge between both types of organization. However, in order for them to keep doing their work, some modifications are proposed in the approach taken by the Hermes Program in filling forms, collecting data, and the frequency and depth with which they work with each one of the schools. In the consultant’s opinion, a few changes over these elements would positively impact the Hermes Program.

On the other hand, with regards to the operation of the program within the schools the perception is not that positive. The first important element is the perception about the perceived commitment of the school’s administrative staff with the implementation of the

program. Administrative staff is widely perceived as being poorly committed with the program for several reasons. First, given that the program does not tackle educational issues related to academic affairs it is seen as not being equally relevant as those types of issues. Second, the program is taken as a means of complying with the Colombian government laws related to ethics, peace, and education in public schools but nothing further than that. Third, stakeholders feel the need of obtaining something from the program, given that it is taking from them (time, work, effort, etc.) instead of rewarding them, they see it more as a burden than as a benefit. For these reasons, the lack of commitment with the program has an impact on the time and space given for conducting the program activities, the perception of the program within and outside the school, the teachers work load, and achieving the voluntary participation of the school's community within the program. Taking this into account, managing the perception of the school's administrative staff turns out to be a key element in improving the operation of the Hermes Program.

Given these results and others broadly presented in the previous sections of this chapter, it is worth noting that the Hermes Program has had a positive impact within the communities in which it operates not only in the students but in the other stakeholders as well. However, in order to produce improvement in the way the program is being implemented, it is important to focus efforts in working with the administrative staff of the school's so that their perception about the program and how they can benefit from it can be positively affected. Given the systemic character of this evaluation, getting to produce a change of the administrative staff perception will have an impact over other activities and issues related to the operation of the program that will positively affect it. It is also important to note that the impact of the program cannot be taken only by exploring how dialogue is used, even if this is the main goal of the program. Significant impacts may be produced in different spheres than those related to the use of dialogue, and by only focusing in this element we may be missing them out. Summing up, significant changes have taken place at a personal level in both students and parents, in improving the relationship among different members of the school community, students have a positive view for their future thanks to the possibilities given by the program in terms of education and skills learned, the ability of identifying inadequate behaviors has increased, and the general perception about the Hermes Program is positive.

The next section will be directed towards exploring additional concepts in the reframed form of Fourth Generation evaluation. In the next chapter I will introduce power issues not only on

the methodological design of the reframed form of Fourth Generation evaluation but in its practice as well.

Chapter 10 – Power issues in the reframed form of Fourth Generation evaluation

Power issues were key in the development of the Fourth Generation evaluation proposed by Guba & Lincoln (1989) as they sought to manage them by means of grounding the evaluation in the formation of constructions, dialogical/hermeneutic circles and negotiation processes. However, I will argue that power could be understood under a different scope, one which does not place it negatively as a defined static characteristic possessed by an individual, but one in constant change. In order to do so, I will ground my work over this issue and its application to my methodological development of Fourth Generation evaluation on the ideas of Michel Foucault.

Why a shift in the underlying power conceptions?

In order to address the traditional conception of power in the program evaluation field, it is important to understand how politics and program evaluation are related. According to Easton (1957), the study of politics is concerned with understanding how authoritative decisions are made and executed for a society and that the output of a political system is a political decision or policy. In bringing such concept to the program evaluation field, Morris (2010) states that the authors of the American Journal of Evaluation (AJE), have approached the notion of politics as the interactions of various stakeholders in evaluation as they articulate their interests from different positions of power, influence, and authority. Taking this into account, Weiss (1993) recognizes program evaluation as taking place in a political context that becomes relevant in three different ways. First, the policies and programs with which the evaluation deals, namely the evaluand, are creatures of political decisions, as they are created as a result of such decisions. Second, because evaluation is undertaken in order to feed into decision making, its reports enter the political arena. Third, evaluation itself has a

political stance. Such political stance comes from its purpose of providing high-quality information to decision makers, automatically claiming a role for itself in the political process, based on the idea that the best information, made available to decision makers would be useful to them (Chelimsky, 1987). Given these definitions, and the recognition of the political character of program evaluation, power emerges as a means by which politics can be enacted. Power is then understood as something that is possessed by someone and which can be applied or inflicted over someone else. Under this assumption, there will always be power differentials among the participants of an evaluation and power will be used “against” others (House, 2015). Taking this into account, power acquires a negative connotation by which someone is coerced, manipulated or imposed something. However, in Guba & Lincoln’s work, power is not only attributed to individuals but to information as well, Guba & Lincoln (1989), state that information is power and evaluation is powerful. The power of an inquiry can be used in a variety of ways inimical to the interests of stakeholder groups. An open determination of who should have what information and when, might lead to a very different resolution of the power issue. However, given that the original form of Fourth Generation evaluation is grounded in the constructivist paradigm, in such circumstances power is shared. Inclusion or exclusion, like most matters in fourth generation evaluation, must be determined by negotiation, in which each audience that wishes to do so can present its case from a position of equal power (Guba & Lincoln , 1989).

Taking into account the arguments related to politics and power in program evaluation and the original form of Fourth Generation evaluation, I would like to approach power issues from a different standing point. The reason for doing this, is particularly related to the methodological characteristics of the reframed form of Fourth Generation methodology. First, the core element of this methodology is directed towards involving the stakeholders of the evaluand throughout the evaluation. In the original design of Fourth Generation evaluation, Lincoln & Guba (1989) argue that negotiations and dialogical circles in the evaluative process will take place with “equal levels of power”. However, I would argue that by conceiving power as being relational and not something that is possessed, it would be easier to understand how power dynamics may produce certain outputs and not others throughout the evaluation stages. In Lincoln & Guba’s argument, power is something that can get to be “manipulated” or handled in a way in which the conditions in which it emerges get to be controlled. However, I consider that instead of handling power one needs to understand it and its implications given that it cannot be eliminated. By comprehending power in terms of

relations, power dynamics between two individuals are not the same as the dynamics between one of those individuals and someone else, this is important given that power does not have a static character but a shifting one that evolves according to the relation between those involved in a situation.

Second, although information may be powerful in the original conception of power in evaluation. I would argue that knowledge and not information is the one which could be attributed as being powerful. This differentiation is important because understanding how knowledge is produced, how certain forms of knowledge are reinforced, how they are given meaning, and how they are used to produce new forms of knowledge is decisive. Third, power has always been understood negatively for which it is seen as a coercive force. I would argue that this does not necessarily happens, given that it can be used in both positive and negative ways by which it can be productive and lead to a better state than the original one. For all these reasons I seek to understand power in terms of the ideas proposed by Michel Foucault with regards to power, knowledge, and truth. These ideas will be described in what follows.

10.1 Foucault on the exercise of power

In order to understand Foucault's conception of power, one first needs to understand what such conception is not. According to Foucault (1990), there are three misconceptions of power embedded in the traditional ways in which the concept is conceived related to its nature, its form, and its unity. As Foucault states:

“By power, I do not mean “Power” as a group of institutions and mechanisms that ensure the subservience of the citizens of a given state [such as characterize many liberal analyses]. By power, I do not mean, either, a mode of subjugation which, in contrast to violence, has the form of the rule [typical of psychoanalytic approaches]. Finally, I do not have in mind a general system of domination exerted by one group over another [i.e. class oppression], a system whose effects, through successive derivations, pervade the entire social body [as in many Marxist views] (1990: 92)”.

Each one of these forms of power may in fact be present in certain contexts as terminal forms, but none are fundamental. Foucault's first task in understanding power is therefore to develop a new method – based on a richer theory – that begins with the basic molecules of power relations and then builds to more complex forms (Lynch, 2011). Such approach opposes the traditional forms in which power is conceived.

Taking this into account, Foucault proposes his own understanding of power as follows.

“It seems to me that power must be understood in the first instance as [1] the multiplicity of force relations immanent in the sphere in which they operate and which constitute their own organization; as [2] the process which, through ceaseless struggles and confrontations, transforms, strengthens, or reverses them; as [3] the support which these force relations find in one another, thus forming a chain or a system, or on the contrary, the disjunctions and contradictions which isolate them from one another; and lastly, as [4] the strategies in which they take effect, whose general design or institutional crystallization is embodied in the state apparatus, in the formulation of the law, in the various social hegemonies (1990: 92–3)”.

One can recognize four important aspects of such definition. The first element corresponds to the multiplicity of force relations operating in a particular domain. Force relations, correspond to whatever in one’s social interactions that pushes, urges or compels one to do something (Lynch, 2011). By understanding power in terms of force relations, it is understood that it develops in specific, local, individual choices, behaviors, and interactions. Such interactions, combine in myriad ways to constitute larger social patterns, and eventually yield macroforms, which one typically thinks of when one thinks about “power” (societies, states, kings) – just as everyday objects are constituted by atoms and molecules (Lynch, 2011). This means that we cannot best understand power by looking at the top of any chain of command. Rather, we must look at the complex webs of interwoven relationships: what Foucault calls the “microphysics” of power (1979: 26). Further, by referring to power in terms of force relations, Foucault draws the attention to the relationship between agents and not to the agent or actor itself. This is particularly important because power dynamics only emerge through the interaction of two agents but not due to the personal traits of those involved in them. Instead,

“...the individual is not a pre-given entity which is seized on by the exercise of power. The individual, with his identity and characteristics, is the product of a relation of power exercised over bodies, multiplicities, movements, desires, forces.” (Foucault, 1980a: 74).

An additional important aspect of Foucault’s definition regarding force relations is the fact that such relations operate in a particular sphere and constitute their own organization. By doing so, particular dynamics in force relations will be produced in specific contexts, dynamics which may not be equally organized in different contexts and will instead acquire a shifting character, meaning that power is not possessed but exercised. And as force relations are “immanent in larger spheres”, they will necessarily be embedded in larger relations as well.

The second element in Foucault’s conception of power corresponds to the fact that force

relations are processes, which instead of being static are constantly being transformed. Such transformations, take the form of confrontations between the original force and its accompanying resistance which may have as a result the strengthening of power relations, its weakening or reversion. Foucault recognizes resistance as a fundamental component in the structure of power, “where there is power, there is resistance, and yet, or rather consequently, this resistance is never in a position of exteriority in relation to power” (1990: 95). Without resistance there is no power relation and through resistance power relations can always be altered. The absence of resistance will lead to relationships of violence and not power. The third element of Foucault’s definition of power corresponds to the interrelationships and systems produced by the processes of the force relations, as various power relations reinforce or undermine each other, by the interactions between power and the resistance to it. The last of the elements recognized in Foucault’s definition of power, corresponds to the strategies built out of the combination of local tactics. At this point, Foucault distinguishes between tactic and strategy. According to Foucault,

“Power relations are both intentional and nonsubjective” (1990: 94). It is intentional in that they are “imbued, through and through, with calculation: there is no power that is exercised without a series of aims and objectives” (ibid: 95).

Such aims and objectives correspond to tactics and strategies both of which are what constitutes rationality,

“the rationality of power is [then] characterized by tactics that are often quite explicit at the restricted level where they are inscribed (the local cynicism of power), tactics which, becoming connected to one another, attracting and propagating one another, but finding their base of support and their condition elsewhere, end by forming comprehensive systems” (ibid.: 94).

Tactics then, correspond to the local rationalities of power in particular cases, whereas strategy corresponds to the larger systemic patterns of power. Conceiving power as strategy, and not as possession means to think of it as something that has to be exerted and not something that can simply be acquired, it is not localized exclusively in certain institutions or individuals, but it is rather a set of relations dispersed throughout society (Balan , 2010).

A fundamental distinction in the work of Foucault on power, is the differentiation between power and violence. In order for a relationship to be consider a power relationship, it must comprise two fundamental elements:

“That 'the other' (the one over whom power is exercised) be thoroughly recognized and maintained to the very end as a person who acts; and that, faced with a relationship of power, a whole field of responses, reactions, results, and possible inventions may open

up” (Dreyfus & Rabinow, 1982).

These two elements correspond to the fact that power recognizes the other as well as involves it in a way in which a response may be produced. On the other hand, violence emerges out of a desire to fix and congeal power relations, by freezing the transfer of power between bodies and then intensifying the stability of whatever 'relation' is left. Because violence attempts to force the will of the one to the will of the other who imposes the acts of violence, it is not brought into play to constitute a subject in difference, but rather because difference is already chosen to be the object of a reductive attack. Such relationship could act upon a body or upon things; closing the door on all possibilities. Its opposite pole can only be passivity, and if it comes up against any resistance it has no other option but to try to minimize it (Barker , 1998). In Foucault's definition of power, resistance is directed towards producing a modification in the action of other(s) but at the same time giving them the possibility of choosing a particular way of reacting. In contrast, violence does not give such liberty to others but instead it seeks to impose a force to a body by which it abstracts from it any possibility of being autonomous. Although both types of relationships are quite different, one can find oneself in a situation with both types of relationships or in one in which power is exercised by means of violence.

In recognizing “the other” and the array of possible responses that can come from them as a way of resistance, “the other” is seen as a free subject.

“Power is exercised only over free subjects, and only insofar as they are free. By this we mean individual or collective subjects who are faced with a field of possibilities in which several ways of behaving, several reactions and diverse comportments may be realized” (Foucault, 1982: 790).

By these means, freedom and its exercise come from the possibility of choosing and not on being imposed a particular course of action. The exercise of power then, “consists in guiding the possibility of conduct and putting in order the possible outcome” (Ibid:789). Such guidance is what Foucault coins as “government”. This idea of governance, does not come from the traditional institutionalized form of imposition but from understanding it as a means in which the possible means of action of others can be structured. Power and freedom do not end up as opposed but in a complex relationship in which they coexist in a reciprocal interaction. If one has the opportunity of choosing between several courses of action or decision making processes, then freedom is present. However, how such decision making process is guided, is led by the emergence of power dynamics. By having the possibility of resistance, power is reshaped and freedom and power dynamics are adjusted for which they

acquire an interactive cyclical relationship that according to Foucault (1982) cannot be separated. “If there are relations of power throughout every social field it is because there is freedom everywhere” (Fornet-Betancourt, Becker, & Gomez-Müller, 1987).

Finally, Foucault does not give power a positive or negative connotation but a productive one. Such point of view is expressed in what follows:

“We must cease once and for all to describe the effects of power in negative terms: it “excludes,” it “represses,” it “censors,” it “abstracts,” it “masks,” it “conceals.” In fact, power produces; it produces reality; it produces domains of objects and rituals of truth. The individual and the knowledge that may be gained of him belong to this production” (1979: 194).

The production of truth by means of knowledge and power will be explored in what follows.

10.2 Power and Knowledge

According to Feder (2011), in order to understand Foucault’s power-knowledge conception, one must first explore the meaning of such concepts. Foucault’s work was conducted in French for which the original words he used may represent different form of knowledge and power which ended up comprised in a single term when translated to English. As such, Foucault (1972) was interested in exploring how implicit knowledge – *savoir* - or the common sense of a particular time, place, or people shaped the explicit knowledge – the *connaissance* – that is institutionalized in the disciplines that make up the human sciences of a particular historical period. On the other hand, the term *pouvoir* translated as power, is also originally used to denote “to be able to”. For this reason, according to Feder (2011) this term should not only be understood in terms of solely referring to power but as a potential capability of doing something. Taking both definitions (knowledge and power) into account, power-knowledge relationship comes to be represented by *pouvoir-savoir* which in turn according to Spivak (1992) corresponds to being able to do something [pouvoir] – only as one is able to make sense of it [savoir]. For Foucault, this knowledge does not have a clear source, it is not the knowledge that is decreed by some authoritative body but one that is "recognized as true", one that can only exist with the support of arrangements of power, those of which do not have a clear origin as no person or body can be said to "have" it (Feder, 2011) for which it is considered to be implicit. Taking into account the adequate language distinctions, I will now move on to explore what is intended by means of the power-knowledge relation.

Foucault initially approaches the relation between power and knowledge in understanding the

change of disciplinary practice over prisoners in the nineteenth century in which the object of punishment shifted from a focus in punishing the body to punishing the soul of the prisoner. Given such transformations, Foucault posed the question regarding which were the causes of such changes as well as how they were produced. In Foucault's (1979) perception, such transformation could be attributed to a shift in the object of discipline as well as a shift in the goals of punishment. A transformation in the object of discipline implied the need of new forms of control and as new dimensions of punishment were being explored, those new dimensions in extend entailed new forms of knowledge. The need for knowledge about the law-breaker who was now identified by his soul more than his individual crime, and the shift of the object of punishment, created new truths about the criminal to be 'discovered' (Hewett, 2004). The criminal is no longer a matter of an individual but of categorizations regarding what is known to be true about the new object of punishment and inquiry. This has an impact on the way the criminal is judged as well, as

“...the sentence that condemns or acquits is not simply a judgment of guilt, a legal decision that lays down punishment; it bears within it an assessment of normality and a technical prescription for a possible normalization” (Foucault, 1979: 20-21).

Normalization for Foucault, corresponds to the institutionalization of the norm, of what counts as normal, indicates the pervasive standards that structure and define social meaning (Feder, 2011). Foucault further explored the concepts of normalization and power-knowledge in his work on the History of Sexuality (1981) in exploring the studies and how sexuality became an important issue in society, how heterosexuality was established as the normal and other sexual conducts were perceived as deviant.

Taking this into account, for Foucault power and knowledge become intimately and productively related. “The exercise of power perpetually creates knowledge and, conversely, knowledge constantly induces effects of power” (Foucault, 1980a). Further, Foucault consider the exploration of such relation not on the basis of a subject of knowledge but on the basis of other three elements as follows.

“These 'power-knowledge relations' are to be analyzed, therefore, not on the basis of a subject of knowledge who is or is not free in relation to the power system, but, on the contrary, the subject who knows, the objects to be known and the modalities of knowledge must be regarded as so many effects of these fundamental implications of power-knowledge and their historical transformations. In short, it is not the activity of the subject of knowledge that produces a corpus of knowledge, useful or resistant to power, but power knowledge, the processes and struggles that traverse it and of which it is made up, that determines the forms and possible domains of knowledge” (1979:27-28).

In considering such statement, truth becomes a historical element which cannot be understood as being absolute or universal. Instead, it acquires a shifting character which is determined by the power relations producing specific forms of knowledge which lead to the establishment of particular forms of truth. 'Truth' is linked in a circular relation with systems of power which produce and sustain it, and to effects" (Fontana & Pasquino, 1979). Attention then is not only drawn to what is considered as being true but to how we take certain things to be true. "'Truth' is to be understood as a system of ordered procedures for the production, regulation, distribution, circulation and operation of statements.

"Truth is a thing of this world: it is produced only by virtue of multiple forms of constraint. And it induces regular effects of power. Each society has its régime of truth, its 'general politics' of truth: that is, the types of discourse which it accepts and makes function as true; the mechanisms and instances which enable one to distinguish true and false statements, the means by which each is sanctioned, the techniques and procedures accorded value in the acquisition of truth; the status of those who are charged with saying what counts as true" (Foucault, 1980a: 131).

Then it becomes impossible for "truth" to acquire an objective character, as it is always influenced by values and principles upon which power relations are grounded. At the same time, for Foucault (1980b) truth should not be understood in terms of it being true or false but over its effects. Foucault does not limit himself only to considering truth as having a political connotation. According to Foucault,

"Everything is political, not because we can take charge of the political and direct everything towards an outcome which we choose, or because we must always take personal responsibility for what occurs in the world, but because the political field is an effect of interacting forces, of the interplay of power, and because relations of force are immanent in the political field" (1990).

As can be seen, the political consideration of Foucault's statement emerges as a result of the relationships in which power is produced.

Given this general framework of the work of Foucault in relation to power and power-knowledge, I will now focus on how power issues in the reframed methodology of Fourth Generation evaluation may be understood under Foucault's ideas.

10.3 Power issues in the reframed form of Fourth Generation evaluation

The relational character of power is present throughout the unfolding of the reframed form of Fourth Generation evaluation not only because the evaluative process evolves, but because in doing so, those involved in it evolve as well. A particular characteristic of the expression of power in this form of evaluation, is that although power is being explored at a micro-level as

Foucault proposes in terms of the considered relations, several of the interactions between those involved in it, represent a relational transference from a larger social scale. Such transference takes place as representatives from identified groups of stakeholders conduct the activities of the evaluation. This representativeness implies that when abstracting individuals from larger groups that follow particular relational dynamics, the general characteristics of such dynamics might be present in the way in which the stakeholders relate within the development of the evaluation. However, this does not mean that relations are based solely on roles attributed to the stakeholder group that is being represented, but to the individual and the individuality of them as well. Both of which are particularly affected by the way in which power relations unfold. For this reason, power relations may not only be explored by means of the relations between individuals but also by exploring their identity by means of the roles they see themselves playing within the evaluand.

For example, power relations are initially present between the client and the evaluator as they are the ones initially involved in the decision making processes over which issues related to the main purpose of the evaluation, and who will be part of it are identified and set. The dynamics between these individuals not only depend on their personal traits but on the role ascribed to each one of them. Traditionally, the evaluator is perceived as the ultimate decision maker, not because of who he is but because of the knowledge and capacities ascribed to the “evaluator” as a role and not as a particular individual. However, in the reframed form of Fourth Generation evaluation, the interactive dynamic between two particular individuals in the roles of client and evaluator will determine how their capacities and knowledge, given their role and their individuality, will be used in directing certain aspects of the evaluation. As the evaluative process moves on, the relational dynamics taken into account in it expand, as more individuals representing different stakeholder groups get involved in the process. At this point, it is worth noting that the power-knowledge ascribed to the client and the evaluator gives birth to who (new stakeholders) will be involved in the evaluation. For example, the evaluator may have a sufficient but not an expert knowledge regarding the evaluand and he may also be an expert how to conduct the evaluation, the client on the other hand, may be an expert over issues related to the evaluand and not the evaluative process, leading to both of them proposing different stakeholders to get involved in further stages of the evaluation. This is, given that they have different perspectives of viewing the same situation. However, the decision making regarding “who” is seen as a relevant stakeholder is taken between both of them at least in an initial stage.

The inclusion of other stakeholders aside of the client, places the evaluator in a pivotal role between the client and the other stakeholders, as they are not directly related but are key decision maker throughout the evaluation. Such pivotal role in the evaluative process, puts the relations involving the evaluator in a key position for directing and completing the evaluative process. According to Foucault (Fornet-Betancourt, 1987), all human relations involve “relationships of power”, i.e., a relation in which someone aims to direct the behavior of someone else. In doing so, the role of the evaluator in playing a pivotal relationship between the stakeholders and the client, turns out to be fundamental in the way in which the evaluation is shaped. The evaluator, acquires an “influencer” role over others involved within the evaluation. On one side, the evaluator is the one responsible of setting the conditions for construction formation by means of dialogical processes involving different stakeholder groups. On the other, the evaluator is in charge of certain decision making processes along with the client. By means of those relationships, information and knowledge considered as relevant emerge, and it is ultimately the evaluator’s decision to establish how further stages of the evaluation will be shaped as he shapes quite actively the field of force relations as well as the issues, claims, and concerns that emerge in them. The way in which the evaluator may be directing the behavior of others is not only evident in prioritizing but also in the way conditions were set. Although the evaluator does not directly influence an individual’s rationality, it may direct its responses by the way in which conditions surrounding decision making processes and construction formation are set, for example by making particular questions and not others, by addressing particular issues of concern while leaving others aside, by sticking to what was initially established between him and the client, among others. However, this does not mean that power is being given a negative connotation. Instead, this means that the created issues, claims, and concerns may be biased by the way in which conditions were set and questions were asked. On the other hand, it is the way in which the generated issues, claims, and concerns are addressed and handled which may acquire a negative or positive character in terms of the validity of the evaluation and the criteria under the process is conducted.

Power also emerges in a different way by involving other stakeholders in decision making processes over the “who” component of the evaluation. Although the evaluator may have a key role in influencing the stakeholders’ responses over particular evaluation issues, the interaction between the stakeholders is as important as the role of the evaluator. Given that the construction formation takes place in a dialogical kind of way, the intervention of

different individuals may have an indirect unrecognized effect over the perceptions of other participants of the dialogical process. For example, an individual may or may not be aware of changes in his own personal perceptions over who may be relevant for the evaluative process given the intervention of someone else. This is to be considered as a power dynamic between two as there it represents a shift in the perceptions of someone given the interaction with someone else.

Further, in order for power to be exercised, resistance necessarily takes place (Foucault, 1990). In Fourth Generation evaluation, resistance is present in the way by which consensus is achieved. As there are several decision making processes taken by different stakeholders, there is a need to achieve some form of consensus over one or several issues in order to proceed with the evaluative process. Resistance has a positive connotation because it is by this means that new constructions are made, new directions for evaluation are taken, and power dynamics between two individuals may be shifted. Resistance implies the possibility of directing not only the relational dynamic between two individuals, and in extent between two stakeholders, in a different direction but also by doing so, directing the issues, claims, and concerns over which there is certain agreement in a different direction than the originally intended or forming new ones which were not originally considered. However, it is worth noting that in a less than ideal situation, forms of coercion, violence, as well as certain forms of biases may be the ones shaping such consensus.

So far I have described how power relations may be present in Fourth Generation evaluation. Another issue related to power relevant to this methodology corresponds to the power-knowledge relation described by Foucault.

10.4 Power-Knowledge and the reframed form of Fourth Generation evaluation

In the reframed form Fourth Generation evaluation, the power-knowledge relation is as important as the concept of power on its own. Power and knowledge are so entangled that Foucault argues that

“There is no power relation without the correlative constitution of a field of knowledge, nor any knowledge that does not presuppose and constitute at the same time power relations” (1979).

This would mean that power relations reinforces particular forms of knowledge not because they may be “truer” but because of the relational dynamics from which they emerge. This may be the case for decision making processes as well as construction formation within this

form of evaluation as they emerge from particular dynamics between the stakeholders of the evaluand.

Throughout the proposed reframed methodology, relationships are considered and reshaped as different stakeholders become part of the evaluative process. Stakeholders not only represent different viewpoints from which the evaluand is being looked from but each one of them also represents different forms of knowledge upon which their perception is grounded. By means of a relational interaction, not only the stakeholders interact but their perceptions as well, and different forms of knowledge are shared so that new constructions get to be formed. Power dynamics in such relations, determine what type of knowledge is perceived as being more valid or useful than others or how new forms of knowledge get to be produced. This does not necessarily mean that unaccounted forms of knowledge are wrong, but that they are not aligned with specific purposes. The knowledge perceived as more valid or adequate for the situation determines future relationships as well, which in turn result in keeping, reshaping or forming new power relations.

Knowledge and what is perceived as true are directly linked. As truth is “a thing of this world”, it does not have an objective universal character. Instead, it has a political connotation based on the veracity attributed to who says that something is true, to particular statements, and what for they are produced. Particular forms of knowledge are reinforced not only because of the power relations that formed them but because of the power that may be attributed to them on their own with respect to the possible outcomes that may be obtained from following particular forms of knowledge. In other words, there are power relations not only between individuals but between different forms of knowledge in producing further and different forms of knowledge as well. This is evident in the reframed form of Fourth Generation Evaluation not only in establishing the purpose with which the evaluation will be conducted, but also in how different forms of knowledge may be helpful in doing so. The initial purpose of the evaluation is defined in a joint manner between the client and the evaluator, this is the first relation in which power dynamics emerge. After an initial purpose has been set, other stakeholders (also identified by the client and the evaluator) are brought in to help shape the “what” component of the evaluation. However, the way in which knowledge related to the second group of stakeholders is handled does not only depends on the relational dynamics of the participants of the evaluation but on the usefulness of such knowledge, how different forms of knowledge may be related, their relationship which the originally intended purpose for the evaluation, and which may be the implications of setting

particular boundaries around them.

In the case of the reframed form of Fourth Generation evaluation, decision making processes regarding boundaries are in the hands of the evaluator. In the previous section I described how the pivotal role of the evaluator between the client and the stakeholders was key in determining the direction given to the evaluation. However, such role is not only important for such relationships but also to the critical role of the evaluator with regards to knowledge. One of the fundamental roles of the evaluator is linked to understanding the implications of the different decision making processes taking places within the evaluation, with regards to the who and the what components. This is shown in the steps of the evaluation methodology that imply setting boundaries and reflecting over them. As Foucault states:

“A critique does not consist in saying that things aren't good the way they are. It consists in seeing on just what type of assumptions, of familiar notions, of established and unexamined ways of thinking the accepted practices are based on... To do criticism is to make harder those acts which are now too easy” (Foucault, 1994).

The understanding of the assumptions, ideas, values upon which certain forms of knowledge are grounded, is significantly useful not only to understand their future implications in framing the evaluation, but also in getting to understand how forms of knowledge (in the form of constructions and setting of objectives) take place within the different decision making processes. Although such situations can no longer be altered, understanding them will lead to correcting mistakes or reshaping the evaluative process if necessary before moving forward with it.

The way in which setting boundaries and reflecting over them takes place is as relevant as the process itself. Although setting boundaries and reflecting over them at least with regards to the “what” component is a decision taken only by the evaluator, it is grounded in guidelines that make it possible to the evaluator to not only rely in his own expertise but on generally approved criteria as well. Such criteria, constitutes an example of what Foucault call “normalization”, as it constitutes a way in grounding judgments with regards to the evaluand in predefined norms. Such norms are the ones that determine what is considered as adequate or useful for conducting future evaluation stages. This leads normalization to having a shaping character over future stages of the evaluation, acquiring a positive character in that it allows evaluation to be more structured in terms of how such processes take place, as they no longer rely on the personal opinion of the evaluator, but at the same time acquiring a negative connotation with respect to missing opportunities to conduct a richer evaluation with issues,

claims, and concerns that may be out of the scope of such norms and that otherwise may have been taken into account. However, this normativity does not imply the dismissal of issues, claims, and concerns, out of the desired scope of what will be included in the evaluation but instead it puts them on hold for a possible future use.

On the other hand, the evaluative process as a whole constitutes a representation of the power-knowledge relation itself. Evaluation is a method of control that combines hierarchical observation with normalizing judgment. For Foucault judgment has a normalizing character because it distinguishes between what is considered as being normal and what is not. An evaluative process as a whole, forms judgments of merit or worth regarding areas of interest of an evaluand. In order to form such judgments, one needs to have expected measures of performance, outcomes, or ideal types of behavior with which real behaviors will be compared. Such expected values are the ones that constitute the normalizing element of evaluation, because by such comparison is that something will be considered as performing adequately (within the norm) or not (the deviant behavior). As a result, by means of the evaluation the behavior and performance of the evaluand or those participating in it comes to be under control. Fourth Generation evaluation is an example of power-knowledge formation, since it combines into a unified whole “the deployment of force and the establishment of truth” (1979: 184). It tells the truth about what or who are being assessed by means of the evaluation, by addressing issues related to the evaluand and its stakeholders and how they are performing, and control their behavior by directing them in a particular direction (Gutting , 2014). Truth also relies upon the chosen methods for addressing the evaluand’s performance, such methods could be qualitative, quantitative, or mixed. In either way, truth will be constructed by means of the information that was gathered and by the meaning given to such findings.

Foucault recognizes three elements as effects of the power-knowledge relation as follows:

“Power relationships carry with them the correlative constitution of a field of knowledge, while knowledge assumes some power relations and at the same time constitutes them. In this scheme, the subject that knows, the various forms of knowledge and the objects to be known must be seen as effects of knowledge-power and the historical transformation thereof” (1989).

The knowledge-power relation in an evaluation process determines the subjects that know, the forms of knowledge, and the object to be known (the evaluand) in the following way. Throughout the evaluation field, the subject that knows shifts according to the focus and the goals with which the evaluation is designed and conducted (evaluation waves). It ranges from

recognizing the evaluator as the unique subject that knows, to recognizing the evaluator and the client, up to recognizing all the stakeholders involved in the evaluation of the program as such. In relation to subjects, the dialogical/hermeneutic processes upon which constructions are discussed, can affect the identity of those who participate in them by redefining their goals in life, the issues, interests, and concerns that interest them. Such modifications in identity and in the individual are a result of the power relations that emerge in the evaluative process. However, although such modifications may be taking place, this does not necessarily mean that those involved in such processes are aware of it.

The second element considered by Foucault as an effect of power-knowledge relations corresponds to different types of knowledge. Different subjects have different forms of knowledge. Although in the reframed form of Fourth Generation evaluation, all subjects considered as relevant stakeholders are placed at the same level, by means of using setting up the conditions for construction formation by means of dialogical processes, not all the knowledge brought to the fore by the stakeholders could be used or taken into account for conducting an evaluation, as this will derive in a lack of rigor of the evaluation and the impossibility of a total comprehensiveness. For this reason, in one way or another prioritization and marginalization of knowledge needs to take place. This takes place by means of drawing boundaries over the subjects (the who component) and the knowledge brought to the evaluation (the what component). When reaching a prioritization and marginalization stage of issues, claims, and concerns, power shifts from the relations to the knowledge under scope. At this point, power-knowledge relations emerge in a more direct manner as knowledge will be the driving force of the evaluation from now on. Initially, the evaluation design depended on who was going to be included in the evaluation by creating boundaries around the stakeholders to be taken into account. Such stakeholders would bring to the evaluation design a particular knowledge influenced by their experiences, ideas, interests, values, and other traits. However, after expressing the issues, claims, and concerns of their interest such knowledge becomes detached from them. From this point on, power shifts from a subject position to the power of knowledge itself. This relation, should be interpreted as considering the possible implications of acknowledging a particular form of knowledge or another as well as what could be further learned from considering such forms of knowledge. In the prioritization case, part of the knowledge is left in a stand by position in which it could be used in further stages of the evaluation or in evaluations conducted in the future.

The third element considered by Foucault in his statement, is the object to be known, which in this case corresponds to the evaluand (that which is going to be evaluated). Deciding what is going to be evaluated is a decision that could be taken by different stakeholders. My methodological design sought to make such decisions in a progressive way.

The initial decision making process is conducted by means of the interaction between the evaluator and the client. They are the participants of this process given that both possess different forms of knowledge relevant for decision making at this stage of the evaluative process that other stakeholders may not have. However, once the decision over where to focus the evaluation has been made, at least in general terms, other stakeholders could be included in helping sharpening such focus. At this point one can identify two different types of knowledge, a general and a specific one. The general knowledge is attributed to the evaluator and the client, and the specific to the other stakeholders. The latter have knowledge that the client will never achieve, as they are part of the program in different ways getting to achieve different perspectives about different issues, claims, and concerns. For this reason, once the general scope of the evaluation has been set, the specific issues, claims, and concerns related to it need to be identified with an active participation of other stakeholders. This point was addressed earlier by recognizing the different subjects that know as well as the activities carried out in both forms of Fourth Generation evaluation for their inclusion and for addressing power issues between them.

When talking about the use of the results of the evaluation, the reciprocal relationship between knowledge and power acquires more relevance. First, knowledge is determinant in giving a particular interpretation to the results of the evaluation, as it directly influences the way in which they are shaped and interpreted by the perspectives over which they are analyzed. Second, the interplay between knowledge and power becomes more important when using the results of the evaluation. The central goal of any evaluative process is the establishment of judgments of merit and worth about an inquiry. However, the way in which these judgments are formulated, determines the possible courses of action that could be taken not only by the evaluator but by the client as well.

10.5 Final Reflections over power issues in the reframed form of Fourth Generation evaluation

The original methodological development of the Fourth Generation evaluation was grounded in a completely different approach to power than the one proposed to explore the reframed

Fourth Generation methodology. Differences in understanding how power operates had an impact on the way in which the methodology was reframed. Originally, power was seen as being possessed by those participating in the evaluation. However, in the approach followed by Foucault which is the one taken into account for the reframed methodology, power is not possessed but instead is a dynamic and shifting relation between those who are being part of the evaluative process. This understanding, gives the opportunity to comprehend how the individuals relate with each other not only at a personal level but as representatives of larger groups of stakeholders as well. As such the dynamics they are embedded in have an impact on their individuality and identity which in extend produces changes in their relational dynamics as well. Such changes are determinant in producing particular forms of knowledge which by means of a prioritization process may become relevant for the evaluation.

One gets to understand that the relational characteristics between two individuals do not necessarily replicate in the relationships between other individuals even if they are surrounded by the same contextual characteristics. This differs from the original conception of Guba & Lincoln in that they considered that by giving all those involved in the evaluation the same opportunities to express their opinions by means of the dialectic/hermeneutic circle, power was being handled. I will argue that power cannot be handled but instead it can be understood by means of understanding the relationships in which it is embedded. By doing so, one can also understand the dynamics by which constructions are formed, as well as understanding how decision making processes of the evaluation take place. The effect of understanding power in this way brought as a result, the inclusion of reflection stages over the formed boundaries regarding the “who” and the “what” components of the evaluation. However, such inclusion was originally based on introducing boundary concepts in the development of evaluation and not in terms of addressing power.

By reflecting over how such decisions were taken, it is possible to understand the power-knowledge dynamic within the evaluation as well. The understanding of power in terms of Foucault’s conception brought as a result, the structuring of norms applied to how prioritization (setting boundaries over the “what”) takes place. The reframed Fourth Generation evaluation included the need to set boundaries around the “who” and the “what” components of the evaluation.

An important aspect in the latter was related to how the boundaries around the “what” component were chosen. Such question poses a major challenge because this type of decision

is traditionally taken as obvious. However, it is grounded on issues related to the role of the stakeholders, power, values and so on. In establishing a more structured ground over which the “how” component stands, transformed prioritization (by establishing boundaries) into a form of a normative process that not only depends on implicit knowledge. By this means, the focus will not only be over the evaluator but over the “power” attributed to the knowledge itself and its capacity to integrate with other forms of knowledge in larger knowledge webs upon which the evaluation will be grounded as well as producing certain outcomes in the form of knowledge (findings) about the evaluand. It is worth noting that this would represent a kind of ideal case given that an ethical component emerges with regards the role of the evaluator and the personal interests that may put pressure over its decision making capacity or that may bias it.

Another important element regarding the conception of power, is the pivotal role ascribed to the evaluator between the client and the other stakeholders of the evaluand. This is a concept which was not originally included in the reframed form of Fourth Generation evaluation but which is very important for its practice. The evaluator is in charge of forming direct relations with both sides of the participants, however it also becomes a bridging figure between them in terms of how by means of such relations the knowledge coming from both sides gets to be handled. This is important because the evaluator itself possesses different forms of knowledge from the ones possessed by other actors, for which power-knowledge relations will determine what is considered as true. The conception over “truth” also differs between the work of Guba & Lincoln, from which the reframed form of Fourth Generation evaluation derives, and the concept of Foucault upon which it is reframed. According to Guba & Lincoln (1989), in the constructivist paradigm “truth” is defined as the more informed and sophisticated construction on which there is consensus among the most competent individuals (not necessarily the most powerful) to form such constructions. However, from the point of view the reframed form of Fourth Generation, evaluation “sophistication”, “competency”, and ascribing someone as “being more powerful” are not decisive in determining what is to be seen as truth. Instead, truth comes from power-knowledge relations, in which particular power dynamics favor the formation of forms of knowledge which result in reinforcing specific truths. Competency and sophistication are no longer relevant in terms that no one is seen as being more competent than someone else in terms of having more knowledge. Instead individuals are seen as possessing different forms of knowledge which in extend determines the formation of diverse constructions. Constructions, grounded in different forms of

knowledge close the door to the possibility of considering a construction as being more sophisticated than other in terms of being properly qualified on its own. Instead, some constructs may be reinforced by means of power relations whereas others may not.

In the reframed form of Fourth Generation evaluation, truth acquires the political connotation given by Foucault. This truth cannot be attributed to have a universal or objective character. Foucault's subjective character of truth goes in hand with what is proposed in the reframed form of Fourth Generation evaluation, in that it is constructed in particular power-knowledge relations that take place at a moment in time. Acknowledging truth in this way, is beneficial for Fourth Generation evaluation by means of understanding that evaluation findings do not constitute an absolute with regards to the evaluand but a particular "picture" of a situation that was drawn over considering certain individuals, stakeholders and methods and by means of the formation of specific relations between them over time. If such process would have been taken in a different period of time or involved different individuals, recognizing different stakeholders (the "who" component), different purposes (the "what" component), or used different methods and the relationships between them, findings may have been completely different.

So far I have addressed power issues in the reframed form of Fourth Generation evaluation and how power-knowledge may be present in this evaluation methodology. The following chapter will address how biases may be present in the theory and practice of this methodology.

Chapter 11 – Biases in the practice and development of the reframed form of Fourth Generation evaluation

The following section will address biases that may emerge in Fourth Generation Evaluation. First, I will give a brief introduction about how biases are produced. Afterwards, I will focus on the biases recognized in the methodology of Fourth Generation evaluation as well as how such biases were handled in practice. Finally, I will address how the consideration of such biases affected the design of the Fourth Generation methodology.

11.1 Study of Biases

Cognitive biases are a component of the research that has been taking part in the last sixty years in behavioral economics, cognitive science, and social psychology with respect to human judgments. Behavioral economics brings to the economy field the idea that not all the human decision making processes have a rational foundation. Instead, economic decisions are influenced by three aspects of human psychology: cognitive limitations (or a bounded rationality), self-control problems, and social preferences (The Royal Swedish Academy of Science, 2017). On the other hand, cognitive science corresponds to the scientific study of the human mind. It is a highly interdisciplinary field, combining ideas and methods from psychology, computer science, linguistics, philosophy, and neuroscience. The broad goal of cognitive science is to characterize the nature of human knowledge – its forms and content – and how that knowledge is used, processed, and acquired (MIT Department of Brain and Cognitive Sciences, 2018). Finally, social psychology has been defined as a branch of psychology that is concerned with those aspects of mental life which relate to social interaction and social phenomena in general (McGarty & Haslam, 1997). Taking into

account, the work on biases has developed in recent years as a way of understanding how human judgments are produced without them necessarily being based on rational patterns of thinking but on other elements related to the interaction of an individual with its context, its personal preferences and appreciations of reality, and how individuals get to be influenced by others. In the next section I will broadly describe the possible sources of biases.

11.2 Sources of Bias

According to Kahneman (2011), cognitive processes can be divided in two conceptual systems: system 1 and system 2. They are used in decision making. System 1 produces intuitive reactions and automatic decisions. It is generated without much conscious effort and channels the available information through a subconscious pattern recognition based on similar past situations (Hogarth, 2001). System 2, corresponds to a deliberate type of thinking, involving focus, reason and analysis. Both systems may be sources of bias. A cognitive bias corresponds to the systematic error in judgment and decision-making common to all human beings which can emerge due to cognitive limitations, motivational factors, and/or adaptations to natural environments (Wilke & Mata , 2012). Systems 1 and 2 cannot be distinguished in real life, however they emerge as a construction of how cognitive processes can be organized. Although both systems are sources of bias, it is more common to find biases emanating from system 1. Putting into practice the use of system 2 by directing decision making processes through logic may be helpful for reducing their appearance. Viewing cognitive processes in terms of these systems makes it possible for us to more easily recognize biases. However this does not directly imply that biases can be completely handled or that they will not emerge again. The human mind has a tendency for being ‘lazy’ for which although the use of system 2 is recommended for reducing the likelihood with which a bias may appear, it is more likely for a human being to use system 1 sometimes resulting in a compulsive decision making process which might not be the most adequate but which is the easiest one.

11.3 Forms of bias in Fourth Generation Evaluation Methodology

Multiple authors have recognized several decision making biases, some of which may be present in the methodology and practice of Fourth Generation evaluation. Since the complete methodological design of Fourth Generation evaluation is explained in detail in Chapter 3. Therefore, I will not refer to it much in what follows.

11.3.1 Common forms of bias

More than 100 cognitive biases have been recognized throughout the literature. However, in what follows I will only address some of the biases that were identified within the Fourth Generation methodology. Other evaluators and practitioners may recognize other forms of bias from the ones described here.

The methodological design of Fourth Generation evaluation relies on the development of constructions by those involved in the evaluative process. Although there are benefits of conducting evaluation in this way, such as getting to involve stakeholders in a more participative manner, it can also produce side-effects that appear as cognitive biases. In the methodological design of Fourth Generation evaluation, constructions are formed at different levels involving within stakeholder group participation, comprising different representatives of the same group of stakeholders, as well as construction formations between stakeholder groups. In both of these processes, biases may emerge.

The first bias to be recognized is the *bandwagon effect* which is a cognitive bias where people will do something simply because other people are doing it, despite their own beliefs, which they may ignore or override. As more people come to believe in something, others also ‘jump on the bandwagon’ regardless of the underlying evidence (Neil , 2016). This bias can be found at both construction formation levels. At first, when referring only to within group construction formation, this type of bias may be present in the form of the identity of the stakeholders. In other words, although all the involved in developing the constructions represent the same group of stakeholders, this does not mean that they have the same issues, claims and concerns of interest. The bandwagon effect may be present if the ideas of someone else are perceived as better as or more useful than the own not because they are better but because of who is asserting them. This bias may also be present in the development of constructions between groups. However, the way in which it emerges is not necessarily linked to the stakeholder’s individual identity but to its role. Both of these dynamics do not emerge at the moment in which constructions are formed but they are transferred from the relationships already present in the evaluand and most likely at higher social spheres as well. Being able to understand such dynamics is not only helpful for reducing the effect of biases like the bandwagon effect but also to understand the shifting power dynamics that may be present in such relationships and interactions. Another way in which the bandwagon bias may be present in the formation of constructions is influenced by the conditions in which such interactions are taking place. For example, a reason for which an individual may choose to

follow other people's ideas may be related to mental tiredness, the haste with which constructions need to be formed, and the amount of time one has been working or focusing on specific tasks. Even though these considerations are not directly linked to the stakeholders' relationships or personal traits, they are equally relevant for the emergence of such bias.

Formulating recommendations also involves the participation of different stakeholders. This is done with the aim of improving the quality of the recommendations as well as the possibility of increasing the likelihood of their application. In this process, the bandwagon effect may appear as well particularly because at this point of the evaluation the issues, claims, and concerns under the scope correspond to the findings of the evaluation and not the particular interests of the stakeholders (including the client). A danger of falling into the trap of the bandwagon bias is that in order to produce a rich construction formation process one needs diversity in thinking, which is the reason why in first place constructions are formed jointly and not individually, if this is not truly achieved, not only the construction formation process but the entire evaluation may diminish its quality.

A similar bias to that of the bandwagon effect is the *courtesy bias*. Williams, Schutt-Ainé, and Cuca (2000) coined the term differential courtesy bias to refer to the situation that would arise if clients found it harder to express negative views for certain types of questions than for others. However, León *et al.*, (2007) extended the concept by raising the possibility that clients may have wanted to please the interviewers in one respect but not another. In other words, this bias is related to the adoption of a more socially correct opinion than the own for fear of offending others. This bias is likely to appear in several stages of the Fourth Generation evaluation methodology. The first case is the one involving the interactions between the evaluator and the client. The relationship between these two stakeholders is a key element as it is initiated in the first stage of the evaluation and continues to be reshaped throughout its stages by serving different purposes. The way in which such relationship is established since the beginning will set the pace for how it evolves throughout further evaluation stages. The role of the evaluator then becomes fundamental in addressing the issues, claims, and concerns of the client. The way in which this bias may emerge in such relationship lies in how the evaluator handles the decisions that need to be made regarding the way in which the evaluation may be conducted. For example, one of the initial decisions that needs to be taken in the evaluation, concern the establishment of the central purpose of the evaluation. Although such decision is supposed to be taken jointly, the evaluator may fall

in the courtesy bias for not being disrespectful with the client's desires and instead of expressing its disagreement with the opinion of the client, it will present his or her point of view in a way in which it may not have any effect over the decision making process, leaving the design and establishment of the evaluation purposes solely to the client or aligning his opinions to that of the client even if they are contradictory. This type of attitude will produce a more submissive relationship between both actors throughout the evaluation as it will place the evaluator in a meek position where it might only be possible for him or her to present results and produce recommendations in a way the client might want to hear but which may not necessarily be the most appropriate one. A second case in which the *courtesy bias* may appear is in the formation of constructions. Once again, in situations involving multiple stakeholders, power issues necessarily emerge. The difference with regards to the consideration of power issues in the bandwagon bias is that in the courtesy bias the counterpart may not be seen in a superior position than the own but instead it is necessarily seen in a different one, whether it is superior or inferior. In such cases, the stakeholder may fall in the courtesy bias by wanting to leave a better impression of himself or by trying not to harm other people feelings.

False consensus effect, a third form of bias, is directly linked to the aforementioned biases as well. Ross *et al.*, define the false consensus effect as people's tendency to "see their own behavioral choices and judgments as relatively common and appropriate to existing circumstances while viewing alternative responses as uncommon, deviant or inappropriate" (1977: p.280). In other words, this bias refers to a process in which people infer that the majority of other people's thoughts, attitudes, and behaviors are aligned with their own (Oostrom , Köbis, Ronay , & Cremers , 2017). This bias may be interpreted in two different way when applied to the Fourth Generation methodology. First, if alternative responses are perceived as "uncommon, deviant or inappropriate" this would imply an imposition of one's own ideas over others. This would mean that a false consensus might be achieved because there is a stakeholder in a dominant position that self imposes its own ideas over others. A situation in which there is no room for carrying out negotiations or dialogical processes in which constructions may be formed but they already come from a single source without leaving room for critique or reshaping them. A second form of false consensus effect may emerge by defining it as the tendency to see the own behavioral choices and judgments as "common". In such form of bias what others have to say is not seen as inappropriate but is taken as similar or equal to what one is saying. Although contrary to what may happen in the

first form of this bias, there is room for dialogue and negotiation. However, such situations may bring as a result the convergence of the speech of the different stakeholders towards what would seem to be the same issues, claims, and concerns. It is in these situations where the bias may emerge. A relevant element for such situation is the used language. Most of the time, when using a complex form of language and having different audiences, may produce the appearance of such bias. If the speaker is not able to communicate its ideas in a way in which all the audiences clearly understand, it is more likely that the false consensus bias may emerge. However, this is not because one perceives other's ideas as similar or inappropriate but because if the audience does not understand, by means of the courtesy bias in which they prefer to agree instead of truly understanding, the false consensus effect may emerge. Specific situations in the Fourth Generation methodology in which the false consensus effect bias may emerge are those in which some form of decision making process involving more than one stakeholder takes place. This is particularly because the false consensus effect emerges as a result of a communication process that necessarily needs a speaker and a receptor.

A fourth form of bias that may emerge in the practice of Fourth Generation evaluation is the *reactive devaluation*. A formal definition of reactive devaluation refers to the process whereby the very act of offering a particular proposal or concession may diminish its apparent value or attractiveness in the eyes of the recipient (Ross, 1995; Ross & Stilling, 1991; Ross & Ward, 1995). In the context of a negotiation, reactive devaluation is likely to lead the parties involved to devalue, and reject, even sincere, good-faith offers presented by the other side, making it harder for the two parties to resolve conflict successfully (Ward, Atkins, Lepper, & Ross, 2011). Although rational inferences may underlie some forms of reactive devaluation (Maoz, Ward, Katz, & Ross, 2002), other mechanisms include significant motivational components— involving changes not in mindless assessments or biased interpretations but rather in the actual preferences of the party receiving a concession offer. Once again, this bias has a relational character which appearance is subject to the communication between two or more individuals. Given that Fourth Generation methodology is based over such interactions, there are particular steps of the methodology in which this bias is likely to appear. As such, the appearance of the bias may appear in three situational contexts: those involving the client appreciation of other stakeholder's viewpoints, the role of the evaluator in prioritizing issues, claims, and concerns of the stakeholders, and construction formation in both; within stakeholder groups, and between stakeholders.

In those situations involving the client, other stakeholders' viewpoints may come to the fore and the way in which the client reacts to such propositions becomes relevant. The appearance of the reactive devaluation bias in this situation, is linked not only to how the client views the other stakeholders but to how he sees himself with respect to them. In this case, reactive devaluation emerges as a result of differences in the hierarchical structure stakeholders are embedded in and how the relationships between them are formed as representatives of stakeholder groups and not only as individuals. In the evaluation field, the client has traditionally seen himself in a superior position when compared to other stakeholders of the evaluand (such as the evaluand's victims and beneficiaries). This view has had an impact in the decision making processes of the evaluation as well as on the development of different evaluation approaches, particularly because other stakeholders aside of the client and the evaluator are not usually seen as valid sources of knowledge. However, the premise over which Fourth Generation evaluation sets its ground, takes stakeholders as constituting a source of different forms of knowledge useful for the evaluand and the evaluation process. Yet, this would not necessarily mean that the reactive devaluation is completely tackled as personality traits, motivations, and perceptions of those in the different stakeholder positions have a direct impact over the emergence of such bias when conducting the evaluation. Although such traits and perceptions may be located at an individual level they are a combination of individual characteristics with social interactions and dynamics.

Another form of bias corresponds to the *in-group bias*. This bias is defined as the tendency to evaluate one's own group more positively in relation to other groups (Aberson , Healy, & Romero , 2000). This bias can be interpreted in three different ways in which it might appear through the Fourth Generation methodology. The first one, is not related towards assessing ones group more positively in terms of qualification but in considering ones group ideas as more valid or adequate than the ones coming from other groups. This does not means that other groups idea are perceived as worst but that the own group ideas are seen as better. In the case of Fourth Generation methodology, this can take place in those situations involving different stakeholders by means of what I would call proximity. This behavior means that individuals are more likely to follow ideas of those whom they consider are more similar to them. This similarity or proximity is not attributed only to belonging to the same group of stakeholders but it may be attributed to other forms of grouping such as gender, age, or even hobbies. In this case if a trait is shared between two individuals, no matter what it is, there may be more affinity between such individuals than that between those who do not have

anything in common. The second way in which in-group biases can take place is by the way one assesses the own group's as well as other group's perception towards certain issues of the evaluand. For example, one may consider other the one's perception with regards to a certain issue as better only because other individuals identify themselves with other groups of stakeholders. In this case an in-group bias emerges because someone is seeing itself as better than someone else not because of the individual traits of such individual but because it is someone that does not belong to the same group. A third form of in-group bias is related to the assessment of the performance of others representatives of different groups of stakeholders when compared to the own. No matter the situation, individuals will always score themselves and those whom with they identify as performing better than others. This is particularly because it is easier to identify other people mistakes than the own, or even recognize that someone is doing better than others.

Curse of Knowledge is a form of bias that can be present not only in Fourth Generation evaluation but throughout the evaluation field. This bias refers to the difficulties one may have for sharing the knowledge we have acquired with others because we cannot readily recreate their state of mind (Heath & Heath, 2006). In other words, when we know something, we find it hard to imagine not knowing it. This implies an impossibility of viewing a situation from the perspective of someone less informed or with a different type of knowledge than us. Curse of knowledge may be present in Fourth Generation evaluation in different ways and through different stakeholders. A way in which this bias could emerge in this form of evaluation, corresponds to those stages involving different stakeholders, for example in construction formation processes and in jointly generating recommendations based on the evaluation findings. In this case, course of knowledge is not only present as a difficulty of one individual in communicating its ideas but in how the audience understands what others are saying. In other words, in negotiation or dialogical processes, it turns out to be very difficult for the audience to put themselves in the position of the speaker in order to understand what it is being said. Given such conditions, conflicts may arise. In this case, the bias appears because different stakeholders have diverse forms of knowledge that are not necessarily superior or inferior to others but just different. Such differences makes it very difficult to understand the point of view from which others are assuming certain positions.

Other forms of biases present in the methodology of Fourth Generation evaluation are *anchoring* and *the focusing effect*. Kahneman, Slovic & Tversky (1982), define the anchoring bias as the action of making estimates by starting from an initial value that is adjusted to yield

the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, different starting points yield different estimates, which are biased toward the initial values. On the other hand, according to Cherubini, Mazzocco & Rumati (2003), the focusing effect bias is related to gathering information about alternatives explicitly stated in a problem context before making a choice without paying attention to other possibilities. In other words, it corresponds to the fact that people make decisions on the basis of the most noticeable and distinct information available in their memory, and for this reason, other pieces of possibly useful information are excluded. So, the decision is only based on fractions of the total amount of knowledge or information that one has stored (WordPress, 2013). Both of these biases are related by the fact that they imply focusing on a narrow set of information for future decision making. Anchoring is particularly related to the use of information for making judgments, whereas the focusing effect is related to causing an error in accurately predicting the utility of a future outcome. In the methodology of Fourth Generation evaluation, such biases may respectively appear in data analysis, in setting boundaries around who and what may be included in the evaluation, and in generating recommendations based on particular findings.

“Confirmation bias”, may also be present in another form as well in the analysis of results. Depending on the use of qualitative or quantitative data, the evaluator must choose the way in which such data may be analyzed and interpreted. Regarding this issue, the confirmation bias emerge when deciding only to use a particular form of data collection and analysis. Doing so may result in confirming the preexistent assumptions of the evaluator by focusing in a single form of analysis. However, different methods such as qualitative and quantitative are used at the same time, cross referring through such results may bring different conclusions and findings than the ones originally expected, such process is called triangulation. By doing so, the chance of having the confirmation bias decreases.

So far I have addressed how traditional forms of bias could be present in the Fourth Generation methodology. The following section will focus on the appearance of other forms out of the ones commonly identified as cognitive biases.

11.3.2 Other forms of bias

Soll, Milkman & Payne (2015), recognize three biases different from the ones described above, that may affect our ability for adequate decision making. Such biases are related to having narrow perspectives regarding the way in which we think about the future, objectives,

and options. Biases related to the way in which we think about objectives, and options are likely to emerge in the Fourth Generation methodology. Each of them will be explained in what follows.

11.4 Biases and the reframed methodology of Fourth Generation Evaluation in Practice

So far I have described how biases may emerge in the reframed methodology of Fourth Generation evaluation. However, I have not addressed how such biases were tackled at a practical level. The case study of this document focuses on evaluating several aspects of a conflict resolution program called Hermes Program. The following section will briefly describe how several of the activities carried out in such evaluation were designed to tackle these biases. So far I have stated that biases may appear in different ways as well as in different stages of the evaluation. For this reason I will describe the steps in which they may take place and what was done to reduce their impact.

In order to start a Fourth Generation evaluation, the evaluator must establish a relationship with the client. When doing so, the initial general purpose with which the evaluation is going to take place will be set as well as the identification of the stakeholders that will be part of the evaluative process. In such partnership, the courtesy bias may appear. A way in which the appearance of such bias could be reduced, depends on the preparation and the personal traits of the evaluator. The first three steps of Fourth Generation evaluation put the evaluator in a position in which he or she needs to gather enough information about the evaluand, how it operates, and the general and specific contexts in which it is immerse. Being able to gather such information gives the evaluator the possibility of forming a rich picture about what is in front of him. This would translate into having the adequate tools and knowledge to face the client at equal positions and being able to actively participate in the process with adequate arguments.

The second phase in which several biases appear, is in developing constructions. Some of them are related to the stakeholders producing such constructions whereas others are related to the constructions as such. In practice, construction formation took place by means of a workshops involving different stakeholders of the program. The first identified bias corresponds to the bandwagon effect. This bias, could surface due to due to the identity of the stakeholders, the role of the stakeholders in the formation of constructions, or by the contextual characteristics surrounding the construction formation process. The way in which such bias was tackled at a practical level was by means of controlling as much as possible the

surrounding conditions of the workshop. In order to do so, adequate space, light fixtures, well defined time for conducting the activities, lunch and coffee breaks, as well as varying the type activities carried out with the stakeholders. By doing so, I tried to reduce the impact of the activities and the extension of the workshop over the ability of the participants in completing the workshop activities.

A second bias in construction formation correspond to the courtesy bias. It was handled by means of a constant reinforcement of the goals with which the stakeholders were being part of the workshop. I sought to reinforce not only the importance of their role within the program but as individuals as well. By doing this, the individual would see himself as being an important part of something bigger, making them more committed to the proposed activities as well as of the importance of expressing what they were really thinking and not what other may have wanted them to say.

On the other hand, a third form of bias called the false consensus effect, was handled through the way in which the activities of the workshop were organized including individual and group activities. In this way triangulation could be made between what participants said individually which the things they said throughout the group activities. This is important because if one only decides to conduct a single type of activity, it becomes more difficult to assess the veracity of the information provided by the participants. By conducting several different activities, 6 in total, it becomes easier to identify behavioral patterns in the stakeholders.

Another bias that may arise in forming constructions corresponds to the reactive devaluation bias between stakeholders. Such bias, was handled by the way in which the small working groups of the workshop were organized. After conducting activities individually, the participants of the workshop were organized in small groups composed by representatives of multiple stakeholder groups. In most of the cases there was at least a representative of each one of the stakeholder groups, without knowing each other. This allowed us to design activities directed towards allowing them to work with one another in a more relaxed way and by forming new relationships. On the other hand, the premise under which construction formation was made was under the “maximum variability sampling” which by gathering different stakeholders would bring a wider variety of viewpoints to the fore, a process which is based on achieved the largest possible variability with respect to the type of information gathered.

One more bias to be identified in construction formation, is the in-group bias. There are two considerations about this bias that need to be taken into account. The first way in which this bias was handled was by the way in which small groups were organized. At most, two members of the same stakeholder group were put together in the same small groups in which most of the activities of the workshop took place. This depended on the number of representatives for each stakeholder group who attended the workshop. Second, although this guaranteed minimizing the impact of the in-group bias with regards to belonging to a particular group of stakeholders, it did not guaranteed that other forms of in-group bias did not emerge. Most of the participants of the workshop were female, for this reason an in-group bias may occur with respect to male or female grouping or another form of grouping based for example on age or the school they belonged to.

The last of the biases found in construction formation is curse of knowledge. This bias is probably the most difficult to handle in construction formation because it would imply producing an impact in the mental constructs and expressed behavior of the stakeholders. However, one of the activities of the workshop based on Edward DeBono six thinking hats sought to manage such bias by means of detaching ideas and knowledge from particular individuals and instead aligning them with particular forms of knowledge represented by the use of different colors.

In prioritizing issues, claims, and concerns several biases emerge as well. The first bias to be handled in such process is the reactive devaluation. This is done by defining standardized criteria for choosing issues, claims, and concerns that will be kept in the evaluation. It is worth noting that although the establishment of such criteria is useful for prioritizing issues, claims, and concerns part of the decision making process relies on the personal opinion of the evaluator or of whoever is in charge of the decision making process making it impossible to completely remove the bias.

The stages related to reflecting over the boundaries of whom and what is considered as part of the evaluation, are two additional stages of the fourth generation methodology which work as mechanisms by which biases can also be tackled. By reflecting over such issues, the evaluator becomes aware of the implications of choosing a particular boundary or another, as well as of the effects this may have in the evaluation. These mechanisms are useful because they constitute a predefined stage of the evaluation and not particular measures within the evaluation steps.

In collecting data one may find a focusing effect bias particularly because the credibility of only using a single data collection form, whether it is qualitative or quantitative, may be put into question. For this reason, the best option would be to use a triangulation mechanism in which several methods are used and which results are further compared. Such methods could be qualitative and quantitative at the same time, with the purpose of achieving a methodological pluralism by which different goals may be achieved and different information needs may be served. By contrasting such results, one may be able to achieve consistency in the findings. After results are analyzed, generating recommendations is the next key step. As this activity also involves the participation of different stakeholders for it to be rich enough, the same biases that present at the construction formation stage will most likely appear. The countermeasures taken in such stage should be replied in this stage as well. However, the difference between both stages lies in the fact that in generating recommendations constructions are not being formed.

The final stage in which biases may appear is in generating reports. In this phase, the curse of knowledge bias may appear as the evaluator will need to adapt to the needs of the different stakeholders for which it will generate reports. The way in which this bias was handled is by first identifying the audiences for which reports will be generated. Afterwards, the evaluator must analyze the particular characteristics of each one of the audiences and those elements that differentiate them from others. Finally, each report produced will need to adapt to such conditions. In this way the curse of knowledge is minimized because the evaluator is not only producing different responses but addressing the needs of those interested as well.

As can be seen throughout the evaluation of the Hermes Program, several measures were taken in reducing as much as possible the effect of the different biases. An important element to be considered in such process is that the stakeholders were not aware of them but instead such measures were naturally incorporated in the way the evaluation was conducted. The following section is directed towards reflecting over the implications of recognizing biases in evaluation and how they have an impact over the methodology and practice of Fourth Generation evaluation.

11.5 Reflections over Fourth Generation evaluation and biases

I have described how biases may be present in the theory and practice of Fourth Generation evaluation. However, the measures taken to address such biases in evaluating the Hermes Program were not directly developed to handle the biases but to handle the constraints

surrounding the evaluation. In addressing such constraints, countermeasures for handling the biases were developed as well, although this was done in an implicit manner. In general terms, every situation involves the appearance of certain forms of biases. However, in Fourth Generation evaluation one can identify two main bias categories: those related to the role of the evaluator and its decision making processes, and those involving other stakeholders as well.

Methodologically, the biases associated solely to the evaluator are related to how decisions regarding the “what” component of the evaluation are made, and to the selection of data collection methods. In order to handle such biases, countermeasures were developed associated with: standardizing how prioritization decisions (“what”) takes place, and choosing more than a single method for data collection in order to achieve triangulation of results. The “how” component of prioritization is linked to a reflective process taking place once boundaries have been set. In this way, the evaluator has the possibility of reshaping the “what” component of the evaluation before moving forward with it by understanding the implications of following an evaluation comprising certain forms of knowledge and not others.

On the other hand, the relational biases that depend on the interaction of two or more stakeholders are more difficult to handle by means of the methodological conception of the evaluation. However, as I have shown so far they can be addressed at a practical level. However, the challenge is not only to address them at a practical level once they appear, but to recognize the possible biases that may emerge before conducting the evaluation. As the methodology of Fourth Generation evaluation has at its core the participation of stakeholders in its different stages, it becomes more important for the evaluator to recognize and deal effectively with such biases.

Maybe the most important lesson for effectively addressing biases in evaluation is that in order to reach different audiences (stakeholders), one needs to know them and not only believe to know them. This implies a real interest in getting to understand their needs, interests, and how to communicate with them, so that the evaluation may be adapted in a way in which they get to be involved in it as more than a source of information for data collection and as stakeholders that matter. This in fact is the ultimate goal with which Fourth Generation evaluation is conducted in the first place.

In the next chapter I will address how the Hermes Program evaluation could be modified.

Chapter 12 –Considerations about the Hermes Program Evaluation

I have already described how the Hermes Program Evaluation was conducted as well as the results obtained from it. In this chapter I seek to address how the evaluation could be improved and how some of its steps could have been conducted in a different way, taking into account that I have already presented power and biases issues in evaluation that I had not previously considered.

12.1 Discussing Results and Generating Recommendations about the Hermes Program

Most of the Hermes Program evaluation was conducted following the methodological design based on the reframed form of Fourth Generation evaluation. However, there are two steps of

the methodology which were not possible to conduct as expected given problems and issues with the operation of the program, as well as strikes and protests affecting the normal operation of Bogota's schools system. For this reason in this section I seek to describe how the activities involving such steps would have been conducted in practice.

Ideally, based on the proposed methodological design, these evaluation stages sought to involve different stakeholders. Given that the stakeholders had previously participated in other evaluation stages, it was expected for them to participate in these stages as well. Although the participation of the same stakeholders was expected, this does not mean that the participation of the same individuals was possible. The first thing to do is to summon representatives of all the relevant stakeholder groups recognized in step 2 of the evaluative process. Given that the initial workshop involving stakeholders included around 40 participants, it is expected to have a similar number of participants for this activity as well. It is important to note, that an ideal characteristic for such group is to have an equal or at least similar number of representatives for each stakeholder group, given that this would increase the possibility of achieving equal participation, at least in quantitative terms, and that the stakeholder representatives may see their participation in the program as equally important. All the activities involving stakeholders are conducted within the premises of one of the Bogota Chamber of Commerce locations. This is done, so that the field may be seen as "neutral" with regards to the program and the different schools in which it operates, and also because the Chamber of Commerce locations, no matter where they are, are very complete in terms of organization, equipment, safety issues, and materials for the workshops.

Once the stakeholder representatives have been summoned, the activity will take place with them and the evaluator. This activity includes three stages: a descriptive, a propositive, and a constructive one. In the descriptive stage it is the role of the evaluator to present how the evaluation was conducted, how the stakeholders and their issues, claims, and concerns were involved in the evaluative process, how data collection was made and which were the findings of the evaluation. This is done with two objectives that are expected to increase the stakeholders' commitment with the evaluation. The first one is related with getting the stakeholders representatives to be aware of how the evaluative process was conducted, this will represent a form of guaranteeing some degree of transparency in the way in which the evaluation was conducted, given that the evaluative process is being explained in detail. Transparency is important in developing the stakeholders' confidence and trust with the evaluative process. Second, it is important for the stakeholders' representatives to see

evidence of their direct or indirect participation in the evaluative process and how their opinions were taken into account not only as a source of information for data collection but for guiding the evaluation as well. These two objectives would make the stakeholders feel more empowered not only with regards to the evaluation but to the evaluand as well, most likely leading to increase their interest in discussing findings and producing recommendations for the future operation of the program as they feel that they and their ideas as useful and taken into account.

Once the descriptive stage has been completed, the second stage of discussing findings and producing recommendations follows. This stage, corresponds to a propositive/proactive stage which is oriented towards discussing findings in a systemic way for which the participants of the workshop are able to propose their ideas about how issues could be tackled and improved. In order to conduct this process in a systemic manner, several issues need to be taken into account. First, the evaluation findings are not to be taken separately, on the contrary several of them must be explored at the same time. This will lead to systemic propositions given that by proposing ideas about how to tackle one finding, other issues may be tackled at the same time, this is less likely to happen if findings are taken separately. Second, in order to propose ideas on how to make improvements or changes over the evaluation findings, the evaluation and the evaluand constraints must be taken into account. In order to do so, the evaluator must present those constraints he or she is aware of to the participants of the workshop. Other constraints from which the evaluator may not be aware of, may emerge as the participants interact as well. The idea of this stage is that participants may propose ideas in a brain storm format. This could be done in two separate stages, the first one in an individual kind of way in which everyone participating in the workshop may be able to participate as pleased. The second stage implies organizing the workshop participants by stakeholder groups. By doing so, they will not only be representing their individual perception about the evaluand, but also by taking into account their affiliation with larger stakeholder groups.

After the propositive/proactive stage, comes the constructive stage. The idea of conducting a stage like this one is related to getting to propose recommendations that are positively valued by the stakeholders and not only by the client or the evaluator. The key element of this stage is getting to address which would be the implications of a recommendation not only over the evaluation results or the evaluand as a whole, but also over those involved in it. For this reason, the ideas that were proposed in the previous step are the ones that will be reshaped, discarded or may be the source of new propositions. The idea is that at the end, the

stakeholders get to dialogically construct recommendations. In order to guarantee that information may not be tampered or manipulated in the future for the client or the evaluator convenience, it may be useful to write and sign and act in which the key outcomes of the discussion are summed up. This is another means by which some degree of transparency in the evaluative process may be achieved.

The recommendations produced in this workshop will then be used to produce the report(s) that will be presented to the different audiences. In order to produce such report, the evaluator as must be aware of the needs and constraints surrounding the audiences and how to fulfill them.

The next section will focus on addressing how the Hermes Program evaluation could have been conducted differently from how it was actually conducted.

12.2 Changes over the Hermes Program Evaluation

In this section I seek to address possible changes in the methodology used to conduct the evaluation of the Hermes Program. I will also address how by considering issues related to biases and power, the used methodology could have been affected.

Before exploring what could have been done differently it is worth brushing up the evaluation methodology. The proposed methodology consists of the following 14 steps: establishing a relationship with the client, understand the context in which the program takes place, identifying adequate stakeholders, reflecting upon the stakeholders' selection, developing, testing, and enlarging constructions, identifying and prioritizing relevant issues, claims, and concerns to be addressed, reflecting upon the boundaries about "what" should be addressed by the evaluation, identification and design of adequate tools for data collection, data collection, analysis of results, discussion and sharing of results to generate recommendation with the stakeholders, generating recommendations, producing reports, and recycling (if needed). Each one of the steps listed about was extensively explored in Chapter 6. In what follows I will address how particular evaluation steps could be conducted differently.

Developing, testing and enlarging constructions

Developing, testing, and enlarging constructions is the step that would be modified the most with regards to the original way in which it was conducted given that it was conducted by means of a workshop with representatives of the relevant stakeholders of the Hermes Program. The following are some things that would have been done differently in the

workshop:

1. The workshop originally included individual activities as well as activities in small groups including representatives from the different stakeholder groups. Aside of these activities the workshop would have included activities in which several individuals of the same stakeholder group would work together before working with other stakeholders' representatives. Organizing activities in this way would be beneficial for the evaluation because it would gather information at different levels: individually, within stakeholder groups, and between groups. Having this opportunity would allow the evaluator to form a richer picture about the perceptions and dynamics within the evaluand and as well as how individuals relate with each other. This is also important in terms of understanding power relations within stakeholder groups particularly because it would lead to understanding how individuals get to form their identity not only at an individual level but also as recognizing themselves as belonging to larger social groups with which they identify. One would be able to explore the characteristics that allow an individual to recognize itself to belonging to one group or to another, and to differentiate from other groups. The three types of interactions will be helpful for understanding and addressing biases, particularly those presented within stakeholder groups. The biases that may arise within stakeholder groups may differ from the ones arising between stakeholder groups as the interaction is primarily based on the individual's identity, personality traits and personal relations, and not to larger social dynamics that could be attributed to playing a role in a relationship with someone who belongs to a different group.
2. Some of the vocabulary used in conducting the activities of the workshop would be modified. An identified issue with regards to the way in which the activities with the stakeholders were conducted, is that the stakeholders tended to understand the same question in very different ways. This was problematic because as the activities required their interaction, understanding the same statement differently, lead the activity in many different directions. This had an effect in the way the participants responded to the questions, or the proposed activities, which in extend led to some difficulties in the interpretation of the results, as well as in getting to form an idea with regards to the issue under the scope.
3. More individuals would be involved in the workshop. Although this was not a difficulty in conducting the first workshop, it would be very useful to involve more

individuals in the different stakeholder groups. Although tentatively the first workshop sought to include at least 5 to 6 representatives of each stakeholder group, for different reasons the workshop did not have an even number of participants representing the different stakeholder groups. For example, there were only 3 students actively involved in the program, whereas there were 7 consultants. Planning the activities considering a larger number of participants would work as a countermeasure for the case in which some of the intended participants could not attend. On the other hand, having more participants would be very helpful in enriching the activities of the workshop and the issues the stakeholders were asked about.

4. Extension of the workshop. The original workshop was conducted in a single session with a 7 hours duration. This was very helpful because activities intended to address different goals were carried out at once. However, the extension of the workshop has an effect over the performance of its participants. Given the extension of the workshop, at the end of the day the participants were very tired as they had been working for several hours. Although their personal commitment in developing the workshop activities was not affected, it could be beneficial for everyone to organize at least two separate shorter sessions for developing, testing, and enlarging constructions. However, I am aware that other constraints may emerge in making such modifications over the design and planning.
5. Incentives. While conducting the first workshop with the stakeholders, it became evident that incentives were an important element for the participants of the workshop. For this reason, if it was to take place again, other forms of incentives would be considered for the stakeholders.

Setting boundaries and reflecting over them

1. Enlarge the evaluation team so that multiple viewpoints enter in discussion over how boundaries may be set and the implications of setting particular ones. The original evaluation design involved a single evaluator (given the constraints surrounding the evaluation). A shortcoming in conducting evaluations with a single evaluator, is that the reflective process is limited to what the evaluator believes is right. While having more staff supporting the evaluator, establishing boundaries and reflecting over them will involve more than one perspective, which in turn may result in better decision making or at least in enriching it. This could be perceived as a way by which power and biases issues may be tackled. It is clear that although the evaluator plays the role

of a facilitator between the stakeholders and the client, at the end, he or she is the one who takes the most important decisions with regards to the evaluation. Expanding the evaluation team would become an opportunity for reinforcing the way by which such decisions are taken as they will no longer rely on a single individual. An effect of this is that biases with regards to a single individual decision making capacity may be limited, although the interaction of different evaluators may bring to the fore other forms of biases as well. However, decision making processes will be taken in a more supported way and by dialogical means which may lead to enriching the decision making process. On the other hand, one of the most important issues in considering power issues is how power-knowledge relations are understood, such relation is important for decision making processes related to the evaluator as power is no longer centered in the relationships between stakeholders but over knowledge.

Data Collection

1. For the Hermes Program evaluation, surveys were used for data collection with regards to consultants and students. However, in collecting data by these means again it would be necessary to look out for other ways by which useful information could be retrieved by means of the survey. Although the surveys were carefully designed by taking into account some of the proposed issues, claims, and concerns of the stakeholders, as well as the central purpose defined with the client, it seems like it could have been better in terms of the collected information. However, this feeling is not related to the design of the questions but to the fact that there were many constraints that limited the space, time, and number of questions for designing and applying the survey. A major constraint was the fact that in order for the students' survey results to be statistically valid in terms of number of observations, more than 350 surveys needed to be answered. This meant that the survey needed to be design in a way in which its answers could be easily computed and interpreted. For this reason, open form questions could not be asked. This was a major restriction given that a person's opinion is not necessarily reflected in answering multiple choice questions. For this reason it is important to consider other ways of designing surveys as well as the possibility of designing other type of activities as well.
2. Try to connect more directly the consultant's findings with those of the students. I have stated that I designed a survey for the students and another for the consultants of the Hermes Program. However, both survey designs differ primarily in that the

consultants' survey was a combination of multiple choice questions with open questions, whereas the students' survey only included multiple choice question. This poses a constraint to the analysis of results given that in at least half of the questions the results given by each stakeholder group were very difficult to connect. For this reason, a modification in developing data collection tools would be related to exploring how both tools could be related not only in analyzing results but also in how they are designed. This would be very beneficial as the systemic character of the evaluation will be reinforced.

3. Widen the scope of the data collection. Although collecting information from the students and the consultants of the program was very significant, given that they are at the center of the program operation, getting information from other stakeholder groups would have been equally useful. Given that all the relevant stakeholder groups were considered in designing the evaluation, it would have been expected that information would have been gathered from all of them as well. However, this did not take place for constraints in both the operation of the program and constraints in developing the evaluation. For this reason, a change in the design of the evaluation will be related to finding ways of including all the relevant stakeholders, or at least some of their representatives, in the data collections stage as well. This could be done by applying surveys or any other data collection tool.

Generating recommendations

1. This stage of the evaluation was not conducted in practice because of internal issues in the operation of the Hermes Program. For this reason, it would have been conducted by the means described in the first section of this chapter.

This chapter has described how some changes could be made to the evaluation of the Hermes Program. I have also described how to carry out the discussion of findings and the generation of results in practice, given that these two stages of the evaluation did not take place. The following chapter is oriented towards presenting the conclusions of the research and how the general and the specific aims were achieved throughout it.

12.3 Power and Biases in the Hermes Program evaluation

Given that power and biases issues have been described after carrying out the proposed methodological design and the Hermes Program evaluation, I will now address globally how power issues and biases may emerge in this evaluation.

First, when exploring power by taking into account Foucault's concepts, the relationship between the evaluator and the client has an effect on the design of data collection tools. The way in which such impact takes place is related to the fact that different interests need to be served and different contextual constraints need to be taken into account while designing the evaluation. The relationship between the client and the evaluator constitutes a way in which power dynamics emerge as a way in which the interests of the client and what they like the most for carrying out the evaluation were considered for data collection. For example, one of the most important interests of the Chamber of Commerce was retrieving and analyzing information by means of quantitative methods. However, the interest of the evaluator were directed towards not only using quantitative methods but also qualitative ones in a linked and back and forth relationship. By these means, the need of producing quantitative results was fulfilled but efforts were extended towards consolidating a broader picture as well. Further stages of the evaluative process, were not only directed towards addressing the client-evaluator relationship but the relationships among the other stakeholders as well as for example, discussing findings and generating recommendations by following a participative approach.

On the other hand, with regards to bias, one can recognize two broad types of biases: individual and group biases. The individual ones were not only related to the partiality of the perspectives of each of the stakeholder representatives but also in the role of the evaluator as decision maker, facilitator, and sometimes without even knowing it advocacy as well. Given that stakeholders relate throughout the stages of the evaluation, relational biases emerge as well, these biases may be explicit or implicit for which the participants may or may not be aware of them. Although the exploration of biases and attempts for handling them may be carried out, this does not guarantee that they can actually be eliminated or their effects get to be removed. Instead, the evaluators and stakeholders are invited to explicitly carrying out the task of identifying them for which the possibility of attempting to address them may be generated.

Chapter 13 – Conclusions

In this chapter I will explain how the aims presented in Chapter 1 were addressed throughout the research. In order to do so, I will present each one of them as well as describing what was done in order to accomplish them.

1. To design, develop and put into action a program evaluation methodology based on the inclusion of systems thinking concepts. Such evaluation approach could be adapted to be used in evaluating social programs involving the participation of multiple stakeholders by taking into account the contextual characteristics surrounding the evaluand, the evaluation and the particular needs that must be fulfilled.

Throughout this research I developed an evaluation methodology based on systems concepts, and particularly those related to boundary critique. Such methodology constitutes a reframing of the Fourth Generation methodology, which is one of numerous evaluation approaches that have been developed so far. In order to do so, I drew upon the work of authors such as Midgley and Ulrich in the systems thinking field (see Chapter 2), the work of Guba & Lincoln on Fourth Generation methodology (Chapter 3), and the work of Cabrera, Colossi & Lobdell and Imam, LaGoy & Williams on introducing systems thinking concepts in evaluation (Chapter 4). However, in order to choose the work of these authors as a guide for my work, a wider review of their work as well as that of other authors in each one of the fields was conducted. This constitutes a contribution to the program evaluation field, at both, theory and practice given that the importance of introducing systems thinking concepts in this field has been recognized and proposed but has not been furthered developed. The contribution itself is divided in two separate sections Chapter 5 and Chapter 6. In Chapter 5, I describe the proposed contribution by addressing why and which concepts were selected, as well as how by combining them, evaluation acquires a systemic character. In Chapter 6, I describe how the evaluation methodology is reframed by describing how the introduction of systems thinking concepts affects the methodological structure by means of modifying, including or eliminating steps. In order to achieve this aim other sub-aims were needed as well.

- 1.1 Reviewing and choosing the most appropriate systems thinking concepts to be included in an evaluation approach.

In order to introduce systems concepts in program evaluation, it was first necessary to review

the most prominent systems thinking approaches that have been developed so far, so that a general understanding on the field could be achieved, and the most appropriate concepts to be introduced in program evaluation could be chosen. For this reason, it became important not only to explore the approaches but to develop a narrative in doing so. In following a narrative, Midgley's proposed narrative was chosen given that it followed a waved structure that made it easy to understand how each wave emerged and evolved from previous ones.

Once the narrative was chosen, the most important developments within each one of the waves were described. Getting to select which were the most important developments, was an issue that not only depended on what I considered as being more relevant but also on which elements were more useful for both the theory and practice of the research. For this reason, the selection of the systems thinking concepts implied a back and forth relationship between the two fields that were going to be considered, namely the systems thinking and the program evaluation fields. Summing up, for each one of the recognized waves, their paradigmatic underlying conceptions were described, as well as their most prominent developments, and the critique that led to the emergence of further waves. Special attention was placed in describing boundary critique, and boundary judgments given that these turned out to be central concepts to be included within the program evaluation field.

1.2 Exploring the work that has been done so far in introducing systems thinking concepts in program evaluation at both, theory and practice.

As I have stated earlier, there have been general attempts on introducing systems thinking concepts in program evaluation. Such developments could be recognized in two different areas, a practical and a theoretical one both of which are described in Chapter 4. By exploring such developments, it is shown that there is a gap in the theoretical and practical work in relating systems thinking concepts in program evaluation. The practical exploration shows that the systems thinking concepts that have been introduced in the practice of program evaluation are directed towards introducing them as tools or methods to conduct a particular step of a methodology but not on grounding the evaluation as a whole. On the other hand, at a theoretical level, systems thinking concepts are largely proposed to be introduced in the program evaluation field, by identifying four important concepts boundaries, perspectives, interrelationships, and the system concept itself. However, there is no further work on illustrating how such concepts may be introduced on a theoretical basis. Such overview raise the possibility of making a contribution in introducing systems thinking concepts at a

theoretical level in a particular evaluation methodology as well as showing how such methodology could be put into practice by conducting an evaluation.

The development of the three sub-aims not only led to the achievement of a larger aim related to producing a methodology and putting it into practice but also in developing an overview of the work that has been done in both fields separately and jointly as well. I recognize that attempts on producing a narrative of both fields separately have been conducted. However, as long as I am aware of, this is the first time that an attempt of addressing both fields separately and jointly is attempted. This is shown in Chapter 4.

1.3 Reviewing and selecting a program evaluation approach (or model) over which systems thinking concepts will be cast upon, an approach which may also be adequate for conducting the practical case.

The main difference in reviewing and selecting relevant systems thinking concepts and appropriate evaluation approaches or models was that the former would be introduced in the latter, which meant that the latter would be chosen with the main intention of being redesigned and not only used. In reviewing and selecting program evaluation approaches, an additional consideration to be taken into account was that theoretically in the program evaluation field, there is a wide arrange of narratives by which the field developments can be accounted for and described. Given this variety, it became important not only to choose a narrative to be followed but to develop one of my own which could be consistent with the one followed in the systems thinking field (this will be described with more detail in aim 4).

Choosing a particular program evaluation model was the most important issue in reviewing the different program evaluation approaches. The criteria used for choosing such method, was based not only in the theoretical foundation of the program evaluation approach, but also on its adequacy for conducting the evaluation of the case study of the research. In reviewing program evaluation approaches, choosing the most relevant was based on criteria related to the purpose with which the evaluation was going to be conducted in practice, as well as exploring the possibilities over which systems concepts could be helpful in reshaping the way in which program evaluations are designed and conducted. In reviewing program evaluation models, their most important features, characteristics, practical cases, and critique were explored. The development of the program evaluation field is presented in Chapter 3.

1.4 Reframing a program evaluation model by taking into account the chosen systems thinking concepts.

Once the work that has been conducted so far in systems thinking, evaluation, and both fields combined had been reviewed, I proceeded to propose a methodological design that included systems thinking concepts in an evaluation model. The description about the chosen concepts and approaches to do so is presented in Chapter 5. This chapter also describes why such concepts were chosen, as well as presenting how considering specific systems thinking concepts will affect key concepts underlying the chosen approach. Once the overall idea about how I propose to incorporate systems thinking concepts in evaluation has been presented, I describe in more detail how the evaluation approach methodological steps are affected by the introduction of systems concepts. This is described in Chapter 6 by presenting the original methodological approach as well as the reframed one, explaining step by step how each one of them is modified or introduced.

2. To understand how power issues may emerge and affect the theory and practice of the proposed methodological design.

Given the political character of evaluation as a whole, power issues are a key issue in conducting evaluations. However, the importance of addressing power in this research does not only lie in exploring power dynamics, but also on understanding how particular conceptions of power may have an effect over how an evaluative process may be conducted. This research constitutes a contribution in addressing power issues in program evaluation particularly because as far as I am aware of, it is the first attempt in addressing power issues from a different perspective from the one traditionally embraced in program evaluation. This research constitutes an attempt of framing power relations in program evaluation based on the conceptions of Michel Foucault's over this issue. Power issues are addressed in Chapter 10. In order to achieve this aim, other sub-aims were to be achieved as well.

- 2.1 Exploring the conceptions of power over which program evaluation is traditionally grounded and particularly the foundations over which the chosen evaluation approach is set.

In order to address power issues under Foucault's conception, it was very important to first understand how power issues were conceived not only in the evaluation field, but also in the chosen evaluation approach. For this reason, I addressed how Guba & Lincoln conceived power issues in their Fourth Generation evaluation approach.

- 2.2 Framing power issues in the chosen evaluation approach in terms of Foucault's conception of power.

Once power conceptions underlying the original methodological development of Fourth Generation evaluation were understood, I decided to explore power based on Foucault's conception in the proposed reframed form of Fourth Generation methodology. In order to do this, I first introduced the concepts developed by Foucault with regards to power, power-knowledge, truth, and violence. Although wider issues considering such conception could have been included, I decided to limit the research to these concepts given the extent of the research and the complexity of Foucault's ideas. Afterwards I proceeded to explore how Foucault's conception differs from the one proposed by Guba & Lincoln in the original Fourth Generation evaluation methodological design as well as exploring how power issues would be present throughout the different stages of the reframed methodology. This itself constituted the most important part of the contribution related to power not only because it was being addressed from a different standing points from its traditional conception in evaluation, but also because evaluation was put under a different scope as well. This may have an impact over how the interaction between stakeholders is conceived, as well as how particular forms of knowledge are reinforced whereas others are discarded in the theory and practice of program evaluation.

2.3 Understand how different power conceptions affect the methodological evaluation development and how this reflects in practice.

Finally, after exploring power issues at a theoretical level, I explored how power issues may have been present throughout the practice of the reframed form of Fourth Generation methodology. I did this by exploring how power issues could be identified and understood in the evaluation of the Hermes Program, which is the case study of this research. However, it is important to note that power issues were addressed after the reframed form of the Fourth Generation evaluation methodology was proposed, and also after the evaluation of the Hermes Program was conducted. The main reason for shifting the power conception over which the evaluation approach was grounded, has to do with the limitations found in the original conception of power proposed by Guba & Lincoln in their original work on Fourth Generation methodology evaluation. Given such limitations, Foucault's proposition over how to conceive power was more in line with the proposed reframed Fourth Generation evaluation methodology and how I wanted to ground it. This is explained in more detail in Chapter 10.

The accomplishment of the sub-aims led to the fulfillment of the larger aim directed towards exploring power issues at both theory and practice in the reframed form of the Fourth

Generation methodology.

3. To explore how cognitive biases may be present in the proposed methodological development and to propose mechanisms by which they could be addressed.

Biases are not a subject that is traditionally explored within the program evaluation field, although in some cases biases are identified in practice. For this reason, getting to explore biases not only in practice but at a theoretical level constituted a contribution to the evaluation field given that up to what I know it is something that has never been attempted. However, it is important to note that the only work I found relating biases and program evaluation is the one conducted by Scriven, such work was conducted more than 40 years ago with a completely different aim and taking into account biases Scriven himself identified. In the last 40 years, research on biases related to human judgment has been widened enough to consider that my attempt of addressing cognitive biases in program evaluation could be regarded as the first one. My attempt of exploring biases is not oriented only towards addressing how they may emerge in the theoretical framework of the reframed form of Fourth Generation evaluation, but also on addressing them in practice, as well as proposing ways by which they could be managed (Chapter 11). In order to make such contribution, the following sub-aims needed to be achieved.

3.1 Review what has been done so far in terms of biases in evaluation.

Given that biases are not commonly handled in the program evaluation field it was important to explore the actual work that has been conducted so far in this subject. The main findings of such review was the fact that the work conducted in relating biases and evaluation is pretty deficient and almost inexistent given that Scriven's work is the only well known in the subject.

3.2 Recognize which are the most relevant cognitive biases for evaluation at both, theory and practice.

In order to recognize which are the most likely biases to emerge in the reframed form of Fourth Generation evaluation methodology, I had to explore the cognitive biases that have been recognized so far and identify those which could emerge and be relevant for its theory and practice. Given that there are around 100 recognized cognitive biases, the current document only includes those which have been identified as being relevant. In doing so, I not only describe how they may emerge at a theoretical and practical level but also what they

mean. This is important because this implied using my own conceptions over biases and the reframed Fourth Generation evaluation methodology as they became intertwined.

3.3 Describe how biases may be present at a theoretical level in the proposed evaluation methodology and how such biases may be handled by the methodological steps and the practice of the evaluation.

The important of describing how biases related to human judgments may be present in the reframed Fourth Generation evaluation was not only on identifying how they may be present at a theoretical level but also on describing how they were handled in practice. This is important because such description presents a series of methods, tools, and practices that were used to address biases. This relevant for the practice of the evaluation field because it not only shows how biases may be theoretically identified in an evaluation methodology, which is itself very important, but also on presenting possibilities for addressing them in practical situations subject to multiple constraints where different stakeholders may be involved. Although my attempt does not constitutes a complete work on this issue, it does constitutes a first step in addressing biases in evaluation, which will open the possibility of conducting more future work over such issues.

4. To develop my own narrative over how the evaluation field has developed so far.

As I have addressed more broadly in the aim 1.2 one of the sub aims related to proposing my own methodology was the need to review how the evaluation field has been developed so far. However, in doing so I encounter the need of developing my own narrative given the need to have consistency between the narrative used to describe the systems thinking field as well as the program evaluation one, and the fact that there are multiple ways by which the evaluation field developments could be described depending on the scope of the used narrative(s). The work related to describing how I designed my own narrative as well as describing the development of the program evaluation field by using it, is described in Chapter 3. The following sub-aims were essential in developing my own program evaluation narrative.

Figure 68. Mapping proposal for describing the evolution of the program evaluation field.

Figure 68 sums up the mapping proposal for describing the development of the program evaluation field. Each one of these waves is differentiated by two elements, who is seen as a relevant decision maker within the evaluative process and upon which type of methods, whether they are quantitative or qualitative, the approaches within the wave rely. Each one of the proposed waves will be described in what follows, as well as the approaches located in each one of them. Particular emphasis is given to describing Fourth Generation Evaluation, located in the fourth wave, as it is the main interest for this research.

4.1 Reviewing different mapping methods developed throughout the evaluation field.

In order to develop my own narrative I first needed to review different ways by which such developments had been described by other authors. A particular characteristic of the evaluation field is that there does not exist a right or wrong way of narrating how the development of the field took place, given that it depends on the elements each one of the narrators chooses to focus its work on. Some of these elements are related to methods, tools, aims, and participants that could be involved in an evaluation approach. For this reason in reviewing mapping forms and developing my own narrative I had to choose elements over which the narrative could be focused.

4.2 Recognizing a differentiating factor between my mapping design and those developed by other authors.

It became relevant to identify traits that allow me to differentiate my mapping work from the mapping methods proposed by other authors, such traits are the ones that make my own narrative a contribution. I did not want to guide my mapping design on a temporally dependant development, although there may be some temporal consistency between the developments, nor on the purpose with which the evaluation was being conducted. Instead, I wanted it to reflect my main interest in conducting evaluation, that is, on exploring how stakeholders are to be included throughout the different evaluation stages. This conception was complemented with one related to how decisions making processes were to take place with regards of the type of tools and information over which they were grounded. Choosing these mapping traits as the ones over which the waves were based, implied a differentiation with respect to the other explored mapping forms given that my own mapping form not only uses different organizing elements but it also organizes the program evaluation approaches differently (although not all of them may differ).

4.3 Recognizing and choosing common traits and patterns over the mapping methods, as well as over the program evaluation approaches.

Given that there needed to be consistency between the narrative followed in both the systems thinking field, as well as the program evaluation field, I chose to follow a waved narrative. However, once the program evaluation waves have been identified, it became important to identify the issues that made a wave to be identified as such, as well as the characteristics of each one of the evaluation approaches that made them belong to a wave and not to another. The elements over which I decided to ground my waved narrative were: the way in which valuing is done in each one of the waves and who is in charge of making such valuing (client, evaluator, client and evaluator, other stakeholders). On the other hand, with regards to the program evaluation approaches, as it happened with the systems thinking field, particular criteria needed to be established in order to choose the program evaluation approaches that were considered as more relevant, given the large amount of approaches that have been developed within this field. The criteria used to make such decision was based on both; theory and practice. The theoretical component was related to the conceptual contribution of the evaluation approach to the program evaluation field, and the practical one was related to the popularity of the approach, exemplified in its degree of recognition and how frequently they were recognized to be used and mentioned in the literature.

4.4 Identifying the issues and critiques that led to further developments in evaluation.

Another important issue over following a waved narrative in the evaluation field is the fact that a movement from a wave to another needed to take place, and as such a reason for such transition needed to be identified. For this reason, another aspect of this narrative was to explore the critique that led to a shift between one wave to another. Given that such waves are not explicitly identified I attempted to establish some form of consistency not only on the criteria used in defining waves but also between such criteria and the critique that led to a shift between one wave and another. This consistency was important in narrative terms given that it added some form of smoothness to the transitions between the waves. It is important to note that critique must not be seen in a negative but instead it could be seen as a creative force, one that directs knowledge and thought in a different direction than the original one leading to new developments and new insights.

5. To conduct the evaluation of a program developed by the Bogota Chamber of Commerce by using the proposed methodology development.

Conducting the evaluation of the Hermes Program constituted the main practical aim of this research not only because it was complementary to the theoretical development of the research but also because it was an opportunity for showing that the proposed theoretical development could be put into practice. In carrying out the evaluation the following sub-aims were to be achieved.

- 5.1 Learning what the Hermes Program does, how it operates, and the contextual factors surrounding it.

Learning about issues related to the Hermes Program constituted the first step in making any attempt at developing a methodological design to evaluate it. Getting to understand such elements constituted a way of framing the evaluation given that the operational characteristics and the contextual factors surrounding the program and the evaluation represented the initial considerations to be taken into account by the evaluation design.

The work of different individuals became crucial in learning and forming the theoretical background of the Hermes Program. Given that the Bogota Chamber of Commerce has a restricted access to the information regarding the operation of the Hermes Program, external documents, theses, working papers, interviews, and others were used to structure the description of its operation. Chapter 7 describes the most important aspects related to the emergence and operation of the Hermes Program.

5.2 Adapting the proposed methodological design to the context and needs in which it will be applied.

Given that the proposed methodology of the research had been proposed and described in detail in Chapters 5 and 6, Chapter 8 described how the methodology was adapted to the Hermes Program. Such adaptation not only described how each one of the proposed steps was conducted but also how customized activities and methods were used in doing so, which were the contextual factors surrounding and impacting the program, as well as who were the stakeholders involved in it. In conducting the evaluation of the Hermes Program methods and tools such as Critical Systems Heuristics (CSH), De Bono's six thinking hats, activities for boosting confidence, drawings, and others were fundamental in conducting the activities with the different stakeholders. Developing such methodology was not only based on introducing systems thinking concepts in evaluation but also on promoting methodological pluralism as well.

5.3 Addressing biases and power issues in the practical case.

Given that power and biases became important concepts in the exploration of the reframed form of Fourth Generation evaluation, it became important to explore such concepts in practice as well. Although I have stated that these concepts are important for the methodological development, they cannot be directly identified in the practical steps of the reframed form of Fourth Generation evaluation for which they need a separate analysis.

The way in which each one of the explicitly defined objectives was achieved throughout the research was explained in detail by means of addressing them individually. However, it is also important to recognize that the way in which qualitative and quantitative methods have been linked turns out to be an additional contribution of the research at both theory and practice. The way in which this link was formed, is exemplified by the Hermes Program evaluation in which designing and using quantitative data collection instruments was grounded on the results of qualitative instruments that were previously designed and used with representatives of the stakeholder groups relevant for the evaluation. And at the same time the quantitative results of using such methods then became the foundation over which qualitative methods were used. The most important element of the relationship between both types of methods is the fact that methodological pluralism can be achieved by mixing methods belonging to both of these types. Using qualitative and quantitative methods within an evaluation approach constitutes a way in which results get to be triangulated to be

reinforced, reshaped or put into question and by which a richer picture of the evaluative process can be achieved. The practical results of doing this are directed towards not only improving the quality of the evaluative process but also towards gaining a deeper insight over the situation that is being studied.

Once the whole evaluation has been conducted, it is important to carry out a reflexive process by which modifications in how the evaluation was conducted and its methodology could be proposed.

References

Aberson , C. L., Healy, M., & Romero , V. (2000). Ingroup Bias and Self-Esteem: A Meta-Analysis.

- Abma , T. (2005). Responsive evaluation: Its meaning and special contribution to health promotion. *Evaluation and Program Planning*, 279-289.
- Ackoff , R. (1974). *Redesigning the Future*. New York : Wiley .
- Ackoff, R. (1979). The Future of Operational Research is Past. *Journal of the Operational Research Society*.
- Adams , F. (2014). Methodological and Epistemic Framework: From Positivism to Post-positivism. En F. Adams, *Measuring National Innovation Performance*, 74 (págs. 1-7). The Innovation Union Scoreboard Revisited.
- Adham, K. A., Kasimin, H., Said, M. F., & Igel , B. (2011). Functions and Inter-Relationships of Operating Agencies in Policy Implementation from a Viable System Perspective. *System Practice and Action Research* , 149-170.
- Advisory, Conciliation and Arbitration Service (ACAS). (2011). *Mediation Explained* . London: ACAS Publications.
- Alkin , M. (2004). *Evaluation Roots: Tracing Theorists' Views and Influences*. Sage Publications.
- Alkin , M. (2012). *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences*. Thousand Oaks, CA: Sage.
- Alkin, M. (2003). Evaluation Utilization Introduction. En T. Kellaghan , D. Stufflebeam , & L. A. Wingate , *International Handbook of Educational Evaluation* (págs. 189-196). Dordrecht: Kluwer Publications .
- American Evaluation Association. (2014). *American Evaluation Association. What is Evaluation?* Obtenido de <http://www.eval.org/p/bl/et/blogid=2&blogaid=4>
- Angyal, A. (1941). A Logic of Systems. En F. E. Emery , *Systems Thinking* (págs. 17-29). Suffolk: Penguin Books.
- Argyris, C., & SchÖn , D. (1978). *Organizational Learning*. London : Addinso-Wesley.
- Army Air Force Center for Low Intensity Conflict. (1989). *Low Intensity Conflict, Overview, Definitions, and Policy Concerns*. Langley Air Force Base, Virginia: United States Department of the Air Force. United States Department of the Army.
- Attemborough, K. (2006). Soft Systems in a Hardening World: Evaluating Urban Regeneration. En B. Williams, & I. Imam, *Systems Concepts in Evaluation An Expert Antology* (págs. 75-89). Inverness: EdgePress.
- Austin , D. (1981). The development of clinic sociology . *Journal of Applied Behavioral Science* , 347-350.
- Balan , S. (2010). M. Foucault's view view on power relations. *Cogito, Vol 2(2)*, 1-6.
- Barker , P. (1998). *Michel Foucault: An Introduction* . Edinburgh: Edinburgh University Press.
- Barnes, M., Matka , E., & Sullivan , H. (2003). Evidence, Understanding and Complexity. *Evaluation* , 265 - 284.
- Barnett, W. S., Escobar , C. M., & Ravsten , M. T. (1988). Parent and clinic early intervention for

- children with language handicaps; a cost-effectiveness analysis. *Journal of the Division for Early Childhood*, 12, 290-298.
- Barone , T. (2000). *Aesthetics, politics, and educational inquiry: Essays and examples* . New York : Peter Lang Publishers.
- Bateson, G. (1972). The logical categories of learning and communication. En G. Bateson, *Steps to an ecology of mind* (págs. 279-308). Chicago: The University of Chicago Press.
- Beer , S. (1972). *Brain of the Firm*. London : Allen Lane.
- Beer, S. (1985). *Diagnosing the System for Organizations*. Chichester: Wiley .
- Biddle , B. J., & Ellena , W. J. (1964). The application of program theory to the evaluation of a managed mental health care system. *Evaluation and Program Planning*, 19, 111-119.
- Boruch , R. (1997). *Randomized experiments for planning and evaluation : a practical guide*. Thousand Oaks; CA: Sage Publications .
- Boruch , R. (2003). Randomized Field Trials in Education. En T. Kellaghan , D. Stufflebeam , & L. A. Wingate , *International Handbook of Educational Evaluation* (págs. 107-121). Dordrecht: Kluwer Academic.
- Boulding, K. E. (1956). General System Theory - The skeleton of science . *Management Science*, 197-208.
- Boyd , A., Geerling , T., Gregory , W. J., Kagan , C., Midgley , G., Murray , P., & Walsh , M. P. (2007). Systemic evaluation: a participative, multi-method approach. *Journal of the Operational Research Society* , 1306-1320.
- Boyd , A., Geerling , T., Gregory , W. J., Kagan , C., Midgley , G., Murray , P., & Walsh , M. P. (2007). Systemic evaluation: a participative, multi-method approach. *Journal of the Operational Research Society*, 1306-1320.
- Brockelsby , J., Cummings , S., & Davies , J. (1995). Demystifying the Viable Systems Model as a Tool For Organisational Analysis. *Asia-Pacific Journal of Operational Research* , 65-86.
- Burghardt, J., & Gordon, A. (1990). *More jobs and higher pay: How an integrated program compares with traditional programs*. New York : Rockefeller Foundation.
- Burke, D. D. (2006). System Dynamics-based Computer Simulations and Evaluation. En B. Williams , & I. Imam, *Systems Concepts in Evaluation* (pág. 226). Inverness: EdgePress.
- Burrell, G., & Morgan, G. (1979). *Sociological Paradigms and Organisational Analysis. Elements of the Sociology of Corporate Life*. Hants, England: Ashgate Publishing Limited.
- Byrne , D. S., & Callaghan , G. (2014). *Complexity Theory and the Social Sciences: The State of the Art*. London : Routledge.
- Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation*, 217 - 228.
- Cabrera, D., Colosi , L., & Lobdell, C. (2008). Systems Thinking. *Evaluation and Program Planning*, 299-310.
- Callaghan , G. (2008). Evaluation and Negotiated Order: Developing the application of complexity theory . *Evaluation* , 399-411.

- Cámara de Comercio de Bogota. (2016). *Cámara de Comercio de Bogotá Lanza Red de Egresados del Programa para la Convivencia Escolar Hermes*. Bogotá: Cámara de Comercio de Bogotá.
- Cámara de Comercio de Bogota. (2017a). *Encuesta de Percepción y Victimización en Bogotá*. Bogotá: Cámara de Comercio de Bogotá.
- Cámara de Comercio de Bogota. (26 de October de 2017c). *Center for Arbitration and Conciliation*. Obtenido de About Us: <http://www.centroarbitrajeconciliacion.com/en/About-us-CAC/About-us>
- Camara de Comercio de Bogota. (2018a, March 21). *What is the Chamber of Commerce of Bogota?* Retrieved from <https://www.ccb.org.co/en/The-Chamber-CCB>
- Cámara de Comercio de Bogotá. (2018b). *Centro de Arbitraje y Conciliación*. Obtenido de What is Arbitration?: <https://www.centroarbitrajeconciliacion.com/en/Services/Domestic-Arbitration/What-is-it>
- Cámara de Comercio de Bogota. (2018c). *Testimonios y Casos de éxito*. Retrieved from <https://www.centroarbitrajeconciliacion.com/Servicios/Convivencia-estudiantil/Testimonios-y-casos-de-exito>
- Cambridge University Press. (2018). *Cambridge Academic Content Dictionary* . Obtenido de Definition of "Generation": <https://dictionary.cambridge.org/us/dictionary/english/generation>
- Campbell , D. T. (1969). Reforms and Experiments. *American Psychologist*, 409-429.
- Campbell , D. T., & Stanley , J. C. (1963). *Eperimental and Quasi-Experimental Designs for Research*. Boston : Houghton Mifflin Company.
- Carden , F., & Alkin, M. (2012). Evaluation Roots: An International Perspective . *Journal of Multidisciplinary Evaluation* , 102-118.
- Carey , T. A., & Stiles , W. B. (2016). Some problems with randomized controlled trials and some viable alternatives. *Clinical Psychology and Psychotherapy*, 23, 87-95.
- Casad, B. J. (2016, August 16). *Encyclopædia Britannica*. Retrieved from Confirmation Bias: <https://www.britannica.com/science/confirmation-bias>
- Centro Nacional de Memoria Historica. (2015). *Una Nacion Desplazada: Informe Nacional del Desplazamiento Forzado en Colombia*. Bogota: CNMH-UARIV.
- Checkland , P. (1980). Are organization machines? . *Futures*, 421-424.
- Checkland , P. (1986). Review of "Diagnosing the system" . *European Journal of Operational Research* , 269-270.
- Checkland, P. (1985). From Optimizing to Learning: A Development of Systems Thinking for the 1990s. *Journal of the Operational Research Society* , 757-767.
- Chelimsky, E. (1987). What have we learned about the politics of program evaluation? *Educational Evaluation and Policy Analysis*, Vol 9, 199-213.
- Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. *American Journal of Evaluation* , 35-55.

- Chen , H. T. (1990). *Theory-driven Evaluations*. Thousand Oaks, CA: Sage Publications .
- Chen , H. T. (2004). The Roots of Theory Driven Evaluation: Current views and origins. En M. Alkin , *Evaluation Roots Tracing Theorists' Views and Influences* (págs. 132-153). Thousand Oaks, CA: Sage Publications .
- Chen , H. T. (2005). *Practical program evaluation: assessing and improving planning, implementation and effectiveness*. Thousand Oaks, CA: Sage Publications.
- Chen , H. T. (2012). Theory-driven evaluation: conceptual framework, application and advancement. En R. Strobl, O. Lobermeier, & W. Heitmeyer , *Evaluation von Programmen und Projekten für eine demokratische Kultur* (págs. 17-40). Wiesbaden: Springer.
- Chen , H., & Rossi, P. (1989). Issues in the Theory-driven Perspective. *Evaluation and Program Planning* , 299-306.
- Chinman , M., Imm, P., & Wandersman , A. (2004). *Promoting Accountability Through Methods and Tools for Planning, Implementation and Evaluation*. Santa Monica: RAND Corporation .
- Churchman , C. W. (1971). *The Design of Inquiring Systems, Basic Concepts of Systems and Organizations*. New York: Basic Books.
- Churchman, C. W. (1970). Operations Research as a Profession. *Management Science*.
- Cochran , W. G. (1983). *Planning and analysis of observational studies* . New York : John Wiley.
- Cohen , C., & Midgley , G. (1994). *The North Humberside Diversion from Custody Project for Mentally Disordered Offenders: Research Report*. Hull : Centre for Systems Studies, University of Hull.
- Cohen , D. (1990). A revolution in one classroom: The case of Mrs Oublier . *Educational Evaluation and Policy Analysis*, 12, 311-329.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cordoba, J. R., & Midgley , G. (2008). Beyond organisational agendas: using boundary critique to facilitate the inclusion of societal concerns in information systems planning. *European Journal of Information Systems* , 125-142.
- Cordoba, J., & Midgley , G. (2006). Broadening the boundaries: an application of critical systems thinking to IS planning in Colombia . *Journal of Operational Research Society* , 1064-1680.
- Cornwall, A. (2008). Unpacking "Participation": models, meanings and practices. *Community Development Journal*, 43, 269-283.
- Cousins , J. B., & Whitmore , E. (1998). Framing Participatory Evaluation . *New Directions for Evaluation* , 5-23.
- Cousins , J. B., Donohue, J. J., & Bloom, G. A. (1996). Collaborative Evaluation in North America: Evaluators' Self-reported Opinions, Practices and Consequences. *American Journal of Evaluation*, 17 , 207-226.
- Cousins, J. B., Whitmore , E., & Shulha, L. (2013). Arguments for a Common Set of Principles for Collaborative Inquiry in Evaluation . *American Journal of Evaluation* , 34, 7-22.
- Cox, P. J., Keener, D., Woodrard, T. L., & Wandersman , A. (2009). *Evaluation for Improvement: A*

Seven-Step Empowerment Evaluation Approach for Violence Prevention Organization .
Atlanta : Centers for Disease Control and Prevention .

Cronbach , L. (1982). *Designing evaluations of educational and social programs*. San Francisco :
Jossey-Bass.

Cronbach , L. (1982). *Designing Evaluations of Educational and Social Programs*. Jossey Bass.

Cronbach, L., Ambron , S. R., Dombusch , S. M., Hess, R. D., Hornik , R. C., Phillips , D. C., . . . Weiner ,
S. (1980). *Toward Reform of Program Evaluation: Aims, Methods, and Institutional
Arrangements*. Jossey-Bass Inc.

Daalgard , B. R., Lewis , D. R., & Boyer, C. M. (1984). Cost and Effectiveness considerations in the use
of computer-assisted instruction in economics. *Journal of Economic Evaluation*, 15, 309-323.

Davies, R., & Dart, J. (2005). *The Most Significant Change (MSC) Technique - A Guide to Its Use*.
United Kingdom : CARE International, United KingdomOxfam Community Aid Abroad,
Australia | Learning to Learn, Government of South AustraliaOxfam New Zealand | Christian
Aid, United Kingdom | Exchange, United Kingdom Ibis, Denmark | Mellemfolkeligt Samvirke
(MS), Denmark.

De Bono, E. (1985). *Six Thinking Hats*. Boston: Little Brown and Company .

de Geus, A. P. (1992). Modelling to predict or to learn? *European Journal of Operational Research* , 1-
5.

Deaton , A., & Cartwright, N. (2017). Understanding and Misunderstanding Randomized Controlled
Trials . *National Bureau of Economic Research* , 1-71.

Dennis , M., & Boruch , R. (1989). Randomized Experiments for Planning and Testing Projects in
Developing Countries. *Evaluation Review*, 292-309.

Departamento Administrativo Nacional de Estadística (DANE). (2017b, March 22). *Pobreza
Monetaria y Multidimensional en Colombia 2016*. Retrieved from
[https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-
vida/pobreza-y-desigualdad/pobreza-monetaria-y-multidimensional-en-colombia-2016](https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-y-desigualdad/pobreza-monetaria-y-multidimensional-en-colombia-2016)

Departamento Administrativo Nacional de Estadística. (2007). *Colombia una Nación Multicultural: Su
diversidad étnica*. Bogotá: Dirección de Censos y Demografía.

Departamento Administrativo Nacional de Estadística. (2017a, June 30). *Reloj de Población*.
Retrieved from <http://www.dane.gov.co/reloj/>

Departamento Administrativo Nacional de Estadística. (2018, January 30). *Gran Encuesta Integrada
de Hogares -GEIH- Mercado Laboral*. Retrieved from
[https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-y-
desempleo](https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-y-desempleo)

Derwisch , S., & Lowe, P. (2015). Systems Dynamics Modelling in Industrial Development Evaluation.
Institute of Development Studies Bulletin , 44-57.

Dick , B. (1993). *You want to do an action research thesis?* Obtenido de
<http://www.scu.edu.au/schools/gcm/ar/art/arthesis.html>

Dick, B. (2000). *Soft systems methodology*. Obtenido de Session 13 of Areol – action research and

evaluation on line.: Available at <http://www.scu.edu.au/schools/gcm/ar/areol/areol-session13.html>

- Dick, B., & Swepton, P. (1994). *Appropriate validity and its attainment within action research: an illustration using soft systems methodology*. Obtenido de <http://www.scu.edu.au/schools/gcm/ar/arp/sofsys2.html>
- Dickson, R., & Saunders, M. (2014). Developmental evaluation: Lessons for evaluative practice from the SEARCH Program. *Evaluation*, 20, 176-194.
- Dispute Resolution Hamburg. (2018). *Conciliation*. Retrieved from What is Conciliation?: <http://www.dispute-resolution-hamburg.com/conciliation/what-is-conciliation/>
- Dozois, E., Langlois, M., & Blanchet-Cohen, N. (2010). *DE 201: A Practitioner's Guide to Developmental Evaluation*. Montreal: The J.W McConnell Family Foundation and the International Institute for Child Rights and Development.
- Dreyfus, H. L., & Rabinow, P. (1982). *Michel Foucault: Beyond Structuralism and Hermeneutics*. Brighton: The Harvest Press.
- Dryzek, J. S. (2000). *Deliberative democratic and beyond*. Oxford: Oxford University Press.
- Dugan, M. (2017). An Empowerment Evaluation of a Comprehensive Sex Education Initiative. En D. Fetterman, L. Rodriguez-Campos, & A. P. Zukoski, *Collaborative; Participatory, and Empowerment Evaluation* (págs. 90-104). New York: Guilford Press.
- Dutt, P. E. (1994). Problem contexts - A consultant's perspective. *Systems Practice*, 539-550.
- Eisner, E. (1976). Educational connoisseurship and criticism: Their form and functions in educational evaluation. *Journal of Aesthetic Evaluation*, 10, 135-150.
- Eisner, E. (1985). *The Art of Educational Evaluation: A Personal View*. Philadelphia: Taylor & Francis Group Inc.
- Eisner, E. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York: Prentice Hall.
- Eisner, E. (1998). *The kind of schools we need: personal essays*. Portsmouth: NH: Heinemann.
- Eisner, E. (2003). Educational Connoisseurship and Educational Criticism: An Arts-Based Approach to Educational Evaluation. En T. Kellaghan, D. Stufflebeam, & L. A. Wingate, *International Handbook of Educational Evaluation* (págs. 153-167). Dordrecht: Kluwer Publishers.
- El Espectador. (2016, March 3). *Al 61,7 % de los colombianos los ingresos apenas les alcanza para subsistir*. Retrieved from <http://www.semana.com/economia/articulo/indice-de-pobreza-monetaria-y-multidimensional-2016-en-colombia-segun-el-dane/463788>
- El Tiempo. (2017, December 27). *Bogotá terminará el 2017 con la tasa de homicidios más baja en 32 años*. Retrieved from <http://www.eltiempo.com/bogota/tasa-de-homicidios-en-bogota-en-el-2017-165462>
- El Tiempo. (2016, June 18). *Cuatro de cada diez escolares han probado drogas*. Retrieved from <http://www.eltiempo.com/archivo/documento/CMS-16624127>
- El Tiempo. (9 de January de 2018). *Deserción escolar bajó a 2,1 %: 6.000 jóvenes volvieron a estudiar*. Obtenido de <http://www.eltiempo.com/bogota/baja-la-desercion-escolar-en-bogota->

- Elster, J. (1998). *Deliberative democracy*. Cambridge : Cambridge University Press.
- Emery , D. (1959). Managerial Leadership through motivation by objectives. *Personnel Psychology* , 65-79.
- Emery , F., & Thorsrud, E. (1976). *Democracy at work*. Astén, Netherlands : H.E Stenfort Kroese .
- Eoyang, G. H. (2006). Human Systems Dynamics: Complexity-based Approach to a Complex Evaluation. En B. Williams, & I. Imam, *Systems Concepts in Evaluation An Expert Antology* (pág. 226). Inverness: EdgePress.
- Fagen, M. C., Redman, S. D., Stacks, J., Barrett, V., Thullen , B., Altenor, S., & Neiger, B. L. (2011). Developmental Evaluation: Building Innovations in Complex Environments . *Evaluation and Practice*, 5, 645-650.
- Feder, E. K. (2011). Power/Knowledge. En D. Taylor, *Michel Foucault: Key Concepts* (págs. 55-68). Durham: Routledge.
- Federal Research Division. (2010). *Colombia: a country study*. Washington D.C: Library of the Congress.
- Feigl , H. (1969). The origin and spirit of logical positivism. En P. Achinstein, & S. F. Baker, *The legacy of logical positivism* (págs. 3-23). Baltimore: Johns Hopkins University Press.
- Fetterman , D., & Ravitz, J. (2017). A Google-Enhanced Empowerment Evaluatio Approach in a GRadute School Program. En D. Fetterman , L. Rodríguez-Campos, & A. P. Zukoski, *Collaborative, Participatory, and Empowerment Evaluation* (págs. 105-115). New York : Guilford Press.
- Fetterman , D., & Wandersman , A. (2005). *Empowerment Evalaution Principles in Practice*. New York : Guilford Press.
- Fetterman , D., & Wandersman , A. (2017). Essentials of Empowerment Evaluation. In D. Fetterman , L. Rodríguez-Campos, & A. P. Zukoski, *Collaborative, Participative, and Empowerment Evaluation Approaches* (pp. 74-89). New York : The Guilford Press .
- Fetterman , D., Rodríguez-Campos, L., Wandersman, A., O'Sullivan, R. G., & Zukoski, A. P. (2017b). An Introduction to Collaborative, Participatory, and Empowerment Evalaution Approaches. En D. Fetterman, L. Rodríguez-Campos, & A. P. Zukoski, *Collaborative, Participatory, and Empowerment Evaluation* (págs. 1-10). New York : Guilford Press.
- Fetterman, D. (2015). Empowerment Evaluation: Theories, principles, concepts, and steps. En D. Fetterman , S. J. Kaftarian , & A. Wandersman , *Empowerment Evaluation: Knowledge and Tools for Self-Assessment, Evaluation Capacity Building, and Accountability* (págs. 20-42). Thousand Oaks: Sage Publications.
- Fetterman, D. M., Kaftarian , S. J., & Wandersman , A. (1996). *Empowerment Evaluation: Knwoledge and Tools for Self-Assessment and Accountability* . Thousand Oaks: Sage Publications.
- Fetterman, D., Rodríguez-Campos , L., Wandersman, A., & O'Sullivan, R. (2014). Collaborative, participatory, and empowerment evaluation: Building a strong conceptual foundation for stakeholder involvement approaches to evaluation (a response to Cousins, Whitmore, and Shulha). *American Journal of Evaluation*, 35, 144-148.

- Fetterman, D., Rodriguez-Campos, L., Wandersman, A., O'Sullivan, R. G., & Zukoski, A. P. (2017a). Similarities Across the Three Approaches: Principles and Practices in Common. In D. Fetterman, L. Rodriguez-Campos, & A. P. Zukoski, *Collaborative, Participatory and Empowerment Evaluation* (pp. 118-135). New York: Guilford Press.
- Fitch, D. (2006). A Cybernetic Evaluation of Organizational Information Systems. En B. Williams, & I. Imam, *Systems Concepts in Evaluation* (pág. 226). Inverness: EdgePress.
- Fitzpatrick, J., Sanders, J., & Worthen, B. (2010). *Program Evaluation: Alternative Approaches and Practical Guidelines*. Pearson.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines*. New York : Longman Inc.
- Fletcher, J. D., Hawley, D. E., & Piele, P. K. (1990). Costs, effects, and utility of microcomputer assisted instruction in the classroom. *American Educational Research Journal*, 27, 783-806.
- Flood, R., & Jackson, M. (1991a). *Critical Systems Thinking: Current research and practice*. John Wiley & Sons Ltd.
- Flood, R. (1995). Total System Intervention (TSI): A reconstruction. *The Journal of the Operational Research Society*, 174-191.
- Flood, R. (1996). Total Systems Intervention: Local Systemic Intervention. En R. Flood, & N. Romm, *Critical Systems Thinking: Current Research and Practice* (págs. 95-117). New York: Plenum Press.
- Flood, R., & Jackson, M. (1991). *Creative Problem Solving: Total System Intervention*. Chichester: John Wiley and Sons.
- Flood, R., & Jackson, M. (1991a). Total Systems Intervention: A practical face to Critical Systems Thinking. *Systems Practice*.
- Flood, R., & Jackson, M. (1991c). *Critical systems thinking: Directed readings*. New York : Wiley & Sons.
- Flood, R., & Jackson, M. (1991b). *Creative Problem Solving Total Systems Intervention*. Wiley & Sons.
- Flood, R., & Romm, N. (1996b). Diversity Management Theory in Action. En Flood Robert, & N. Romm, *Critical Systems Thinking*: (págs. 81-93). New York : Plenum Press.
- Flood, R., & Ulrich, W. (1990). Testament to conversations on critical systems thinking between two systems practitioners. *Systems Practice*, 7-29.
- Fluhr, J. D., Oman, R. F., Allen, J. R., Lanphier, M. G., & McLeroy, K. R. (2004). A Collaborative Approach to Program Evaluation of Community Based Teen Pregnancy Prevention Projects. *Health Promotion Practice*, 5, 127-137.
- Fontana, A., & Pasquino, P. (1979). Truth and Power: An Interview with Michel Foucault. *Critique of Anthropology*, Vol 4, 131-137.
- Foot, J. L., Gregor, J. E., Hepi, M. C., Baker, V. E., Houston, D. J., & Midgley, G. (2007). Systemic problem structuring applied to community involvement in water conservation. *Journal of the Operational Research Society*, 645-654.
- Fornet-Betancourt, R., Becker, H., & Gomez-Müller, A. (1987). The ethic of care for the self as a

- practice of freedom: an interview with Michel Foucault on January 20, 1984. *Philosophy and Social Criticism*, 12, 112-131.
- Forrester, J. (1962). Industrial Dynamics. *Science*, 426-427.
- Forss, K., Marra, M., & Schwartz, R. (2011). *Evaluating the Complex: Attribution, Contribution and Beyond*. Transaction Publishers.
- Foucault, M. (1976). Two Lectures. En C. Gordon, *Michel Foucault. Power/Knowledge* (págs. 78-108). London: Harvester Wheatsheaf.
- Foucault, M. (1980). Power/Knowledge. En C. Gordon, *Truth and Power* (pág. 114). New York: Pantheon Books.
- Foucault, M. (1980a). Power/Knowledge. En C. Gordon, *Truth and Power* (pág. 114). New York: Pantheon Books.
- Foucault, M. (1990). *The History of Sexuality, Volume I: An Introduction*. New York: Vintage.
- Foucault, M. (1990a). *The History of Sexuality, Volume I: An Introduction*. New York: Vintage.
- Foucault, M. (1994). En J. Faubion, *Power Essential works of Michel Foucault, 1954-1984* (pág. 456). London: Penguin.
- Fredericks, K., Deegan, M., & Carman, J. (2008). Using Systems Dynamics as an Evaluation Tool. Experience from a Demonstration Program. *American Journal of Evaluation*, 251-267.
- Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach*. Cambridge: Cambridge University Press.
- Fuller, B., Hua, H., & Snyder, C. W. (1994). When girls learn more than boys: The influence of time in school and pedagogy in Botswana. *Comparative Education Review*, 38, 347-376.
- Gamble, J. (2008). *A Developmental Evaluation Primer*. Canada: The J.W. McConnell Family Foundation.
- Gates, E. (2016). Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning*, 62-73.
- Geiß, R. (2006). Asymmetric conflict structures. *International Review of the Red Cross*, 757-777.
- Glaser, B. (1978). *Theoretical Sensitivity*. Mill Valley, CA: Sociology Press.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.
- Glouberman, S., & Zimmerman, B. (2002). *Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like?* Commission on the Future of Health Care in Canada.
- Greene, J. (2000). Challenges in practicing deliberative democratic evaluation. En K. E. Ryan, & L. DeStefano, *Evaluation as a democratic process: Promoting inclusion, dialogue and deliberation* (págs. 13-26). New Directions for Evaluation.
- Gregory, A. (1997). Evaluation Practice and the Tricky Issue of Coercive Contexts. *Systems Practice*, 589-609.
- Gregory, A. (2000). Problematizing Participation A Critical Review of Approaches to Participation in

- Evaluation Theory . *Evaluation* , 179-199.
- Gregory , A., & Jackson, M. (1992). Evaluation Methodologies: A Systems for Use. *The Journal of the Operational Research Society*, 19-28.
- Gregory , W. (1996). Dealing with diversity. En R. Flood, & N. Romm, *Critical Systems Thinking* (págs. 37-62). New York: Plenum Press.
- Gregory , W. J. (1992). *Critical systems thinking and pluralism: a new constellation*. London : PhD Thesis, City University.
- Gregory , W. J. (1996). Discordant pluralism: a new strategy for critical systems thinking . *Systems Practice* , 605-625.
- Guba , A., & Lincoln , Y. (1994). Competing paradigms in qualitative research. En N. K. Denzin , & Y. Lincoln , *Handbook of qualitative research* (págs. 105-117). London : Sage.
- Guba , E., & Lincoln , Y. (1981). *Effective Evaluation* . Michigan : Jossey-Bass Publishers.
- Guba , E., & Lincoln , Y. (1981). *Effective Evaluation* . San Francisco; CA: Jossey-Bass.
- Guba , E., & Lincoln , Y. (1989). *Fourth Generation Evaluation* . Newbury Park: Sage Publications.
- Guba , E., & Lincoln , Y. (1989). *Fourth Generation Evaluation* . Sage Publications .
- Guba , E., & Lincoln , Y. (2001). *GUIDELINES AND CHECKLIST (a.k.a FOURTH GENERATION) EVALUATION* . Kalamazoo: West Michigan University .
- Gueron , J. M., & Pauly , E. (1991). *From welfare to work* . New York : Russell Sage Foundation.
- Gutmann , A., & Thompson , D. (1996). *Democracy and Disagreement* . Cambridge, MA: Belknap Press.
- Gutting , G. (2014). Michel Foucault . *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University .
- Hakan , K., & Seval , F. (2011). CIPP evaluation model scale: development, reliability and validity . *Procedial Social Behavioral Sciences*, 592-599.
- Hargreaves , M. B., & Podems , D. (2012). Advancing Systems Thinking in Evaluation: A Review of Four Publications. *American Journal of Evaluation*, 33, 462-470.
- Harré, R. (1981). *Great Scientific Experiments* . Oxford, UK : Phaidon Press.
- Hart , D., & Paucar-Caceres, A. (2017). A utilisation focussed and viable systems approach for evaluating technology supported learning. *European Journal of Operational Research* , 626-641.
- Haslett , T., & Sarah , R. (2006). Using the Viable Systems Model to Structure a System Dynamics Mapping and Modeling Project for the Australian Taxation Office. *Systems Practice and Action Research* , 273-290.
- Heath, C., & Heath, D. (2006, December). *The Curse of Knowledge*. Retrieved from Harvard Business Review: <https://hbr.org/2006/12/the-curse-of-knowledge>
- Hewett, M. A. (2004). Michel Foucault: power/knowledge and epistemological prescriptions. *University of Richmond Scholarship Repository*, 2-30.

- Hogarth, R. M. (2001). *Educating Intuition* . Chicago: University of Chicago Press.
- Hordijk, L. (2007). The art and craft of Systems Analysis. *Options Magazine*, 14-15.
- Hou Tay , B., & Kee Pong Lim, B. (2006). Using Dialectic Soft Systems Methodology as an Ongoing Self-evaluation Process for a Singapore Railway Service Provider. En B. Williams , & I. Imam , *Systems Concepts in Evaluation An Expert Antology* (págs. 89-101). Inverness: EdgePress.
- House , E. (1983). Assumptions underlying evaluation models. En G. Madaus , M. Scriven , & D. Stufflebeam , *Evaluatio Models Viewpoints on Educational and Human Services Evaluation* . The Hague : Kluwer-Nijhoff Publishing.
- House , E. (1997). Evaluation in the Government Marketplace. *American Journal of Evaluation* , 37-48.
- House , E., & Howe , K. (2003). Deliberative Democratic Evaluation. En T. Kellaghan , D. Stufflebeam, & L. Wingate, *International Handbook of Educational Evaluation* (págs. 79-102). Dordrecht: Kluwer Academic.
- House, E. (2015). The politics of evaluation and the evaluation of politics. *Evaluation, Vol 21*, 492-496.
- Howe , E., & Howe , K. (1999). *Values in Evaluation and Social Research* . Thousand Oaks, CA: Sage Publications.
- Hummelbrunner, R. (2011). Systems Thinking and Evaluation. *Evaluation, Vol 17(4)*, 395-403.
- Hummelbrunner, R. (2011). Systems Thinking and evaluation . *Evaluation* , 395-403.
- Hummelbrunner, R., & Reynolds, M. (2013). Systems thinking, learning and values in evaluation . *Evaluation COnnections: The European Evaluation Society Newsletter*, 9-12.
- Imam, I., LaGoy, A., & Williams, B. (2006). Introduction. En B. Williams, & I. Imam, *Systems Concepts in Evaluation* (págs. 3-10). Inverness: EdgePress.
- Ireh , M. (2016). *Scientific Management Still Endures in Education* . Winston Salem, NC: Winston Salem State University.
- Jachimowicz, J. (2015 , March 25). *Huffington Post* . Retrieved from https://www.huffingtonpost.com/jon-jachimowicz/randomized-controlled-tri_b_6507928.html
- Jackson , M. (1988). An appreciation of Stafford Beer's "Viable System" viewpoint on managerial practice. *Journal of Management Studies* , 557-573.
- Jackson , M. (1982). The Nature of Soft Systems Thinking: The work of Churchman, Ackoff and Checkland. *Journal of Applied Systems Analysis*, 17-28.
- Jackson , M. (1987). New Directions in Management Science. En M. Jackson , & P. Keys, *New Directions in Management Science* (págs. 133-159). Aldershot : Gower .
- Jackson , M. (1987). Present Positions and Future Prospects in Management Science. *Omega*, 455-466.
- Jackson , M. (1991a). Modernism, Post-Modernism and Contemporary Systems Thinking. En R. Flood, & M. Jackson, *Critical Systems Thinking: Directed Readings* (págs. 287-303). Willey &

Sons.

- Jackson , M. (1991b). *Systems Methodology for the Management Sciences*. New York: Plenum Press.
- Jackson , M. (2000). *Systems Approaches to Management*. New York : Plenum.
- Jackson , M. (2003). *Systems Thinking: Creative Holism for Managers*. Chichester: John Wiley & Sons .
- Jackson, M. (1985). Social Systems Theory and Practice: The Need for a Critical Approach. *International Journal of General Systems*.
- Jackson, M. (1999). Towards coherent pluralism in management science . *Journal of Operational Research Society* , 12-22.
- Jackson, M. (2001). Critical Systems Thinking and practice . *European Journal of Operational Research* , 233-244.
- Jenkins, G. M. (1972). The systems approach. *Systems Behaviour*, 78-104.
- Jenkins, G. M., & Youle, P. V. (1968). A Systems Approach to Management. *Operational Research Society*, 5-21.
- Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology Research and Development*, 5-14 .
- Karlsson, O. (1996). A critical dialogue in evaluation: How can interaction between evaluation and politics be tackled? *Evaluation* , 405-416.
- Katz , J., & Gartner , W. (1988). Properties of emerging organizations . *Academy of Management Review* , 429-441.
- Kliebard , H. (1970). Reappraisal: The Tyler Rationale. *The School Review*, 78, 259-272.
- Kozol, J. (1991). *Savage inequalities: Children in America's schools*. New York : Trumpet Club .
- Kramer , J. B. (2015). The Place of Eisner's Educational Connoisseurship and Criticism in Jewish Education. *International Journal of Jewish Education Research* , 8, 67-79.
- Kruglanski, A. W., & Kroy , M. (1976). Outcome validity in experimental research: A reconceptualization . *Journal of Representative Research in Social Psychology* , 7 , 168-178.
- Kuhn , T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kushner, S. (1992). *A musical education: innovation in the Conservatoire*. Victoria, Australia: Deakin University Press.
- Lam , C. Y., & Shulha, L. M. (2014). Insights on Using Developmental Evaluation for Innovating: A Case Study on the Cocreation of an Innovative Program . *American Journal of Evaluation*, 36, 358-374.
- Laughlin , R., & Broadbent, J. (1996). Redesigning Fourth generation Evaluation An Evaluation Model for the Public-sector Reforms in the UK? *Evaluation*, Vol 2, 431-451.
- Law Reform Commission . (2010). *Alternative Dispute Resolution: Mediation and Conciliation* . Dublin : Law Reform Commission .
- Lax , D. A., & Sebenius, J. K. (2006). *3-D Negotiation: Powerful Tools to Change the Game in Your Most Important Deals*. Boston : Harvard Business School Press.

- Leonard , A. (2009). The Viable System Model and its Application to Complex Organizations . *Systems Practice and Action Research* , 223-233.
- Levin , H. (1983). *CLO Cost Effectiveness: A Primer*. Michigan : Sage Publications.
- Levin , H., & McEwan , P. (2003). Cost-Effectiveness Analysis as an Evaluation Tool. En T. Kellaghan , D. Stufflebeam , & L. A. Wingate, *International Handbook of Educational Evaluation* (págs. 125-152). Dordrecht: Kluwer Publications.
- Lewis , D. R. (1989). Use of cost-utility decision models in business education . *Journal of Education for Business*, 64, 275-278.
- Lewis , D. R., & Kallsen , L. A. (1995). Multiattribute evaluations: An aid in reallocation decisions in higher education. *The Review of Higher Education*, 18, 437-465.
- Lewis , D. R., Johnson , D. R., Erickson , R. N., & Bruninks, R. H. (1994). Multiattribute evaluation of program alternatives within special education. *Journal of Disability Policy Studies*, 5, 77-112.
- Lincoln , Y., & Guba , E. (1980). The Distinction Between Merit and Worth in Evaluation . *Educational Evaluation and Policy Analysis*, 2, No. 4, 61-71.
- Lincoln , Y., & Guba , E. (1986). Research, evaluation and policy analysis: heuristics and disciplined inquiry. *Policy Studies Review*, 546-565.
- Lynch, R. A. (2011). Foucault's Theory of Power. En D. Taylor, *Michel Foucault: Key Concepts* (págs. 13-27). Durham: Routledge.
- MacDonald , B., & Kushner, S. (1982). *Bread and dreams*. Norwich : University of East Anglia.
- MacNeil, C. (2000). Surfacing the Realpolitik: Democratic evaluation in an antidemocratic climate. En K. E. Ryan , & L. DeStefano , *Evaluation as a democratic process: Promoting inclusion, dialogue and deliberation* (págs. 51-62). New Directions For Evaluation .
- Madaus , G., & Stufflebeam, D. (1989). *Educational Evaluation: Classic Works of Ralph W. Tyler*. Boston: Kluwer Academic Publishers.
- Madaus, G., Scriven , M., & Stufflebeam , D. (1983). *Evaluation Models Viewpoints on Educational and Human Services Evaluation* . Norwell, MA: Kluwer Nijhoff Publishing .
- Mark , M., Greene, j., & Shaw , I. (2006). The Evaluation of Policies, Programs and Practices. En I. Shaw, J. Greene , & M. Mark , *The Sage Handbook of Evaluation* (págs. 1-30). Thousand Oaks: Sage.
- Martin , E., & MacDonald, R. (2012). *A System Dynamics-Based Evaluation of the New York State HIV Testing Law*. New York: University of Albany.
- McGarty, C., & Haslam , S. A. (1997). Introduction and a short history of social philosophy. En C. McGarty, & S. A. Haslam, *The Message of Social Psychology* (págs. 1-19). Oxford: Blackwell.
- McKee, A., & Watts, M. (2000). *Protecting Space? The Case of Practice and Professional Development Plans*. Norwich : Centre for Applied Research in Education, University of East Anglia.
- McMahon , W. W. (1998). Conceptual framework for the analysis of the social benefits of lifelong learning. *Education Economics*, 6, 309-346.
- Mertens , D. M. (2015). Philosophical Assumptions and Program Evaluation . *Spazio Filosofico*, 75-85.

- Midgley , G. (1989). Critical Systems: The Theory and Practice of Partitioning Methodologies. *Proceedings of the 33rd Annual Meeting of the International Society for the Systems Sciences*. Edinburgh: International Society for the Systems Sciences.
- Midgley , G. (1990). Creative Methodology Design. *Systemist*, 108-113.
- Midgley , G. (1990). Creative Methodology Design. *Systemist Vol 12*, 108-113.
- Midgley , G. (1992). Systems Practice Pluralism and the Legitimation of Systems Science. *Systems Practice* , 147-173.
- Midgley , G. (2000). *Systemic Intervention: Philosophy, methodology and practice* . New York: Springer.
- Midgley , G. (2003). Systems Thinking: An introduction and Overview. En G. Midgley , *Systems Thinking, Volume 1: General Systems Theory, Cybernetics and Complexity* (págs. xvii-liii). London : Sage Publications.
- Midgley , G. (2008). Systems thinking, complexity and the philosophy of science. *Emergence: Complexity and Organization* , 55-73.
- Midgley , G. (2011). Theoretical pluralism in Systemic Action Research . *Systems Practice*, 1-15.
- Midgley , G. (2015). Systemic Intervention. En H. Bradbury , *The SAGE Handbook of Action Research 3rd Edition*. London : Sage.
- Midgley , G., & Floyd , M. (1990). Vocational Training in the Use of New Technologies for People with Disabilities. *Behaviour and Information Technology* , 409-424.
- Midgley , G., & Pinzón , L. (2011). Boundary Critique and its implications for conflict prevention. *Journal of the Operational Research Society*, 1543-1554.
- Midgley , G., & Shen, C. (2007). Toward a Buddhist Systems Methodology 2: An exploratory, questioning approach . *Systems Practice* , 195-210.
- Midgley , G., Ahuriri-Driscoll, A., Foote , J., Hepi , M., Taimona , H., Rogers-Koroheke, M., . . . Wood, D. (2007). Practitioner Identity in Systemic Intervention: Reflections on the Promotion of Environmental Health Through Māori Community Development. *Systems Research and Behavioral Science*, 233-247.
- Midgley , G., Munlo , I., & Brown , M. (1998). The Theory and Practice of Boundary Critique: Developing Housing Services for Older People. *Journal of the Operational Research Society*, 467-478.
- Midgley, G. (1996a). What is this thing called CST? En R. Flood, & N. Romm, *Critical Systems Thinking* (págs. 11-24).
- Miller, R. L. (2016). On Messes, Systems Thinking and Evaluation: A Response to Patton . *American Journal of Evaluation*, 37, 266-269.
- Mingers , J. C. (1980). Towards an appropriate social theory for applied systems thinking: critical theory and soft systems methodology. *Journal of Applied Systems Analysis*, 41-50.
- Mingers, J. (1984). Subjectivism and Soft Systems Methodology—A critique . *Journal of Applied Systems Analysis* , 85-101.

- Mingers, J., & Brocklesby, J. (1995). *Multimethodology: towards a framework for mixing methodologies*. Coventry : University of Warwick.
- Miser, H. J., & Quade, E. S. (1985). *Handbook of Systems Analysis*. New York: John Wiley & Sons.
- MIT Department of Brain and Cognitive Sciences. (2018, April 12). *Cognitive Science*. Retrieved from <https://bcs.mit.edu/research/cognitive-science>
- Morell, J. (1984). Making Evaluation Viable: The Response of Graduate Programs in Evaluation. *American Journal of Community Psychology*, 209-2016.
- Moreno, A. (14 de October de 2014). ¿Por qué es mayor la pobreza rural que la urbana? *Portafolio*. Obtenido de <http://www.portafolio.co/economia/finanzas/mayor-pobreza-rural-urbana-62730>
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 113-127.
- Neil, J. (2016, November 9). *Cognitive Bias: The Bandwagon Effect*. Retrieved from The Ethics Centre - Business and Professional Ethics: <http://www.ethics.org.au/on-ethics/blog/november-2016/cognitive-bias-the-bandwagon-effect>
- Newman, J., Martin, L., Velasco, M., & Fantini, A. (2003). A Systems Dynamics Approach to Monitoring and Evaluation at the country level: An Application to the Evaluation of Malaria control Programs in Bolivia. *Fifth Biennial World Bank Conference on Evaluation and Development* (págs. 1-16). Washington : OED Conference.
- Nunnally, R. (2001). The world of evaluation in an imperfect democracy. *Minnesota Evaluation Conference*.
- O'Sullivan, R. (2004). *Practising evaluation: A Collaborative Approach*. Thousand Oaks: Sage Publications.
- Observatorio de Desarrollo Económico. (2016). *Pobreza y Desigualdad en Bogotá: resultados en los últimos años*. Bogotá: Alcaldía Mayor de Bogotá.
- Oostrom, J. K., Köbis, N. C., Ronay, R., & Cremers, M. (2017). False consensus in situational judgment tests: What would others do? *Journal of Research in Personality*, Vol 71, 33-45.
- Orr, L. L., Bloom, H. S., Bell, S. H., Doolittle, F., Lin, W., & Cave, G. (1996). *Does job training for the disadvantaged work? Evidence from the National JTPA Study*. Washington D.C: Urban Institute Press.
- Oxford Dictionary. (March de 2017). *English Oxford Living Dictionaries*. Obtenido de English Oxford Living Dictionaries: <https://en.oxforddictionaries.com/definition/genus>
- Oxford Dictionary. (March de 2017). *English Oxford Living Dictionaries*. Obtenido de English Oxford Living Dictionaries: <https://en.oxforddictionaries.com/definition/stabilization>
- Patton, M. Q. (1978). *Utilization-Focused Evaluation*. Thousand Oaks: Sage Publications.
- Patton, M. Q. (1986). *Utilization-Focused Evaluation*. Newbury Park: Sage Publications.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation: The new century text*. Thousand Oaks: Sage.
- Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. Thousand Oaks; CA: Sage

Publications.

- Patton , M. Q. (2003). Utilization-Focused Evaluation. En T. Kellaghan , D. Stufflebeam , & L. A. Wingate , *International Handbook of Education Evaluation* (págs. 223-244). Dordrecht : Kluwer Publishers.
- Patton , M. Q. (2006). Evaluation for the Way we Work. *Nonprofit Quarterly* , 28-33.
- Patton , M. Q. (26 de July de 2010). *AEA 365 - A Tip-A-Day by and for Evaluators*. Obtenido de Michael Quinn Patton on Developmental Evaluation: <http://aea365.org/blog/michael-quinn-patton-on-developmental-evaluation-applying-complexity-concepts-to-enhance-innovation-and-use/>
- Patton , M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York : The Guildford Press .
- Patton , M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Thousand Oaks: Sage.
- Patton , M. Q. (2012). A utilization-focused approach to contribution analysis. *Evaluation*, 18, 364-377.
- Patton , M. Q. (2012). *Essentials of utilization-focused evaluation* . Thousand Oaks: Sage Publications.
- Patton , M. Q. (2016a). What is Essential in Developmental Evaluation? On Integrity, Fidelity, Adultery, Abstinence, Impotence, Long-Term Commitment, Integrity, and Sensitivity in Implementing Evaluation Models. *American Journal of Evaluation*, 250-265.
- Patton , M. Q. (2016b). The developmental evaluation mindset: Eight guiding principles. En M. Q. Patton , K. McKegg, & N. Wehipeihana, *Developmental evaluation exemplars: Principles in practice* (págs. 289-312). New York : Guilford.
- Patton , M. Q., Asibey , E., Dean-Coffey, J., Kelley , R., Miranda , R., Parker , S., & Newman, G. (28 de January de 2014). *American Evaluation Association* . Obtenido de American Evaluation Association: <http://www.eval.org/p/bl/et/blogid=2&blogaid=4>
- Patton, M. Q. (2012). *Essentials of Utilization Focused Evaluation* . Thousand Oaks: Sage Publications.
- Patton, M. Q. (2013). *Utilization-Focused Evaluation (U-FE) Checklist*. Michigan : Evaluation Checklist Project.
- Polanyi , M. (1964). *The tacit dimension*. Garden City; NY: Doubleday.
- Polanyi, M. (1966). *Personal knowledge Towards a Post-Critical Philosophy* . Chicago : University of Chicago Press .
- Psacharopoulos, G. (1994). Returns to investment in education: A global update . *World Development*, 22, 1325-1343.
- Quade, E. S., & Boucher , W. I. (1968). *Systems Analysis and Policy Planning: Applications in Defense*. Santa Monica, California: The RAND Corporation.
- Ramirez, R., & Brodhead, D. (2013). *Utilization Focused Evaluation: A primer for evaluators*. Penang, Malaysia: Southbound.
- Reynolds , M. (2006). Evaluation Based on Critical Systems Heuristics. En B. Williams, & I. Imam,

- Systems Concepts In Evaluation* (págs. 101-121). Inverness: EdgePress.
- Reynolds , M. (2010). Evaluation and stakeholding development. *The 9th European Evaluation Society International Conference* . Prague, Czech Republic.
- Reynolds , M., Forss , K., Hummelbrunner, R., Marra , M., & Perrin , B. (2012). Complexity, systems thinking and evaluation. *Evaluation Connections Newsletter of the European Evaluation Society*, 7-9.
- Reynolds, M. (2007). Framing purposeful evaluation through critical systems thinking. *Rural Development for the XXIIInd Congress of the European Society for Rural Sociology* (págs. 20-24). Wageningen : Netherlands.
- Reynolds, M., Forss, K., Hummelbrunner, R., Marra , M., & Perrin , B. (2012). Complexity, systems thinking and evaluation - an emerging relationship? *Evaluation Connections: The European Evaluation Newsletter*, 7-9.
- Rodriguez-Campos , L., & Rincones-Gomez , R. (2012). *Collaborative Evaluations: Step-by-Step* . Stanford: Stanford University Press.
- Rodriguez-Campos, L. (2017). Essentials of Collaborative Evaluation. En D. Fetterman , L. Rodriguez-Campos, & A. P. Zukoski, *Collaborative, Participatory, and Empowerment Evaluation* (págs. 10-21). New York : Guilford Press.
- Rogers , P. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation* , 29-48.
- Rossi, P. (2004b). My views of evaluation and their origins. En M. Alkin , & C. Christie, *Evaluation Roots Tracing Theorists' Views and Influences* . Thousand Oaks; CA: Sage Publications .
- Rossi, P. H., Lipsey , M. W., & Freeman, H. E. (2004a). *Evaluation: A Systematic Approach*. Thousand Oaks: Sage Publications.
- Ryan , K. E., & Johnson , T. D. (2000). Democratizing evaluation: meanings and methods from practice. En K. E. Ryan , & L. DeStefano , *Evaluation as a democratic process: Promoting inclusion, dialogue and deliberation* (págs. 39-50). New Directions for Evaluation .
- Sanderson , I. (2000). Evaluation in Complex Policy Systems. *Evaluation* , 433 - 454.
- Sankaran , S., & Tay , B. H. (2003). Action Research Models in Business Research. *Proceedings 9th ANZSYS Conference*. Melbourne.
- Sankaran , S., Tay , B. H., & Cheah, Y. S. (2003). Application of a Dialectical Model of Soft Systems Methodology to Conduct Action Research. *Proceedings, ALARPM/SCIAR Conference*. Queensland.
- Schechter, D. (1991). Critical Systems Thinking in the 1980s: A connective summary. En R. Flood, & Michael Jackson, *Critical Systems Thinking: Directed Readings* (págs. 213-227). Willey & Sons.
- Schnoes , C. J., Murphy-Berman, V., & Chambers, J. M. (2000). Empowerment Evaluation Applied: Experiences, Analysis, and Recommendations from a Case Study. *American Journal of Evaluation* , 53-64.
- Scriven , M. (1966). *Primary Philosophy* . McGraw Hill.
- Scriven , M. (1967). The Methodology of Evaluation. En R. Tyler , R. Gagne , & M. Scriven ,

- Perspectives on curriculum evaluation* (págs. 39-83). Skokie: AERA Monograph Series on Curriculum Evaluation .
- Scriven , M. (1968). An introduction to metaevaluation . *Educational Products Report*, 36-38.
- Scriven , M. (1972). Objectivity and subjectivity in educational research. En L. G. Thomas , *Philosophical redirection of educational research: The seventy-first yearbook of the National Society for the Study of Education* (págs. 94-142). Chicago : University of Chicago Press.
- Scriven , M. (1973). Goal-free Evaluation. En E. R. House, *School evaluation: The politics and process* (págs. 319-328). Berkeley, CA: McCutchan .
- Scriven , M. (1976). Evaluation Bias and its Control. En G. V. Glass, *Evaluation studies review annual* (págs. 101-118). Beverly Hills, CA: Sage.
- Scriven , M. (1980). *The logic of evaluation* . Inverness, CA: Edgepress.
- Scriven , M. (1981). Product Evaluation. En N. L. Smith , *New techniques for evaluation* (págs. 121-166). Beverly Hills, CA: Sage.
- Scriven , M. (1991). *Evaluation Thesaurus* . Thousand Oaks: Sage.
- Scriven, M. (1983). Evaluation ideologies. En G. F. Madaus, M. Scriven, & D. L. Stufflebeam, *Evaluation models: viewpoints on educational and human services evaluation* (págs. 229-260). Boston : kluwer-Nijhoff.
- SEBoK authors. (27 de March de 2017). *Cite the SEBoK*. Obtenido de BKCASE Editorial Board, Guide to the Systems Engineering Body of Knowledge (SEBoK):
http://sebokwiki.org/w/index.php?title=Special:Cite&page=Cite_the_SEBoK&id=52693
- Secretaria de Educación del Distrito. (2017). *Perfil de Cobertura Local 2017*. Bogota: Secretaria de Educación del Distrito Capital.
- Secretaria Distrital de Integración Social. (2017, June 21). *Al día se registran en promedio 76 denuncias por violencia intrafamiliar en Bogota*. Retrieved from
<http://www.integracionsocial.gov.co/index.php/noticias/95-noticias-familia/1962-al-dia-se-registran-en-promedio-76-denuncias-por-violencia-intrafamiliar-en-bogota>
- Shadish , W. R., Cook , T. D., & Leviton , L. C. (1991). *Foundation of Program Evaluation: Theories of Practice* . Newbury Park : Sage Publications.
- Shadish , W., Cook , T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Sharma , S., & Gutierrez , J. A. (2010). An evaluation framework for viable business models for m-commerce in the information technology sector. *Systems Practice and Action Research* , 33-52.
- Shaw , I. F., Greene, J. C., & Mark, M. M. (s.f.).
- Snyder , S. (2013). The Simple, the Complicated, and the Complex: Educational Reform Through the Lens of Complexity Theory. *Organisation for Economic Co-operation and Development*, 1-35.
- Soll, J. B., Milkman, K. L., & Payne , J. W. (2015). Outsmart Your Own Biases. *Harvard Business Review*, 65-71.

- Stake , R. (2000). A modest commitment to the promotion of democracy. En K. E. Ryan , & L. DeStefano , *Evaluation as a democratic process: Promoting inclusion, dialogue and deliberation* (págs. 97-106). New Directions in Evaluation .
- Stake , R. (2003). Responsive Evaluation. En T. Kellaghan , D. Stufflebeam , & L. Wingate , *International Handbook of Educational Evaluation* (págs. 63-68). Dordrecht: Kluwer Academic.
- Stake , R., Migotsky, C., Davis, R., Cisneros , E., Depaul , G., Dunbar, C., . . . Chaves, I. (1997). The evolving synthesis of program value. *Evaluation Practice* , 89-103.
- Stake, R. (1967). The Countenance of Educational Evaluation. *Teachers College Record*, 68, 523-540.
- Stame , N. (2004). Thory-based Evaluation and Types of Complexity. *Evaluation*, 58-76.
- Stanford Encyclopedia of Philosophy . (2013, October 7). *Pragmatism*. Retrieved from <https://plato.stanford.edu/entries/pragmatism/>
- Sterman , J. (2001). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston : McGraw Hill.
- Sterman , J. D. (2000). *Business Dynamics> Systems Thinking and Modeling for a Complex World* . United States : McGraw Hill .
- Stern , E. (2004). *The foundations of evaluation and impact research* . Luxemburg : Office for Official Publications of the European Communities .
- Stoller, A. (2015). Taylorism and the logic of learning outcomes . *Journal of Curriculum Studies*, 47, 317-333.
- Stufflebeam , D. (1983). The CIPP Model for Program Evaluation. En G. F. Madaus , & D. Stufflebeam , *Evaluation in Education and Human Services*, 6. Dordrecht: Springer.
- Stufflebeam , D. (2001). Evaluation Models. *New Directions in Evaluation* .
- Stufflebeam , D. (2003). The CIPP Model for Evaluation. In T. Kellaghan , D. Stufflebeam , & L. A. Wingate, *International Handbook of Educational Evaluation* (pp. 31-62). Dordrecht: Kluwer Publishers.
- Stufflebeam , D. (2004). The 21st Century CIPP Model: Origins, Development, and Use. In M. Alkin, *Evaluation Roots: Tracing Theorists Views and Influences* (pp. 225-260). Sage Publications.
- Stufflebeam , D. (17 de March de 2007). *CIPP Evaluation Model Checklist. A tool for applying the CIPP model to assess long-term enterprises*. Obtenido de West Michigan University : https://www.wmich.edu/sites/default/files/attachments/u350/2014/cippchecklist_mar07.pdf
- Stufflebeam , D. L., & Shinkfield, A. J. (2007). *Evaluation Theory, Models and Applications*. San Francisco: John Wiley & Sons.
- Stufflebeam , D., Madaus , G. F., & Kellaghan , T. (2000). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* . Boston : Kluwer Academic .
- Suchman , E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. New York: Russell Sage Foundation .

- Suescún , J. I., & Jiménez, C. E. (2011). Desplazamiento forzado de los grupos étnicos en Colombia. En F. Del Popolo , E. M. Garcia de Pinto da Cunha, B. Ribotta, & M. Azevedo, *Pueblos indígenas y afrodescendientes en América Latina: dinámicas poblacionales diversas y desafíos comunes* (págs. 127-150). Rio De Janeiro: UNFPA .
- Susman, G. (1976). *Autonomy at work*. New York : Praeger.
- Tay , B. H., & Lim , K. P. (2004). A Scenario-based Training System for a Railway Service Provider in Singapore. *Proceedings, Systems Engineering/ Test and Evaluation Conference 2004*. Aledaide.
- Tepe , S., & Haslett, T. (2002). Occupational health and safety systems, corporate governance and viable systems diagnosis: An action research approach. *Systemic Practice and Action Research* , 509-522.
- The Oxford Dictionary . (September de 2017). *English Oxford Living Dictionaries* . Obtenido de <https://en.oxforddictionaries.com/definition/apperception>
- The Royal Swedish Academy of Science. (2017). *Richard H Thaler: Integrating economics with psychology*. Stockholm : The Royal Swedish Academy of Sciences.
- Thomas, A., & Lockett, M. (1979). Marxism and Systems Research: Values in Practical Action. *Proceedings of the 23rd Annual Conference of the International Society for General Systems Research*.
- Thompson , J. (1967). *Organization in Action*. New York : McGraw Hill .
- Tokmak, H. S., Baturay , M., & Fadde, P. (2013). Applying the Context, Input, Process, Product Evaluation Model for Evaluation, Research and Redesign of an Online Master's Program. *The International Review of Research in Open and Distributed Learning*, 14.
- Torres , R., Preskill, H., & Piontek , M. E. (1996). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations*. Thousand Oaks : Sage Publications.
- Trist, E., & Bamforth, K. (1951). Some social and Psychological Consequences of the Longwall Method of Coal-Getting. *Human Relations*, 3-38.
- Trist, E., Higgin , G., Murray , H., & Pollock, A. (1963). The Assumption of Ordinariness as a Denial Mechanism. Innovation and Conflict in a Coal Mine. En E. Trist , G. Higgin , H. Murray , & A. Pollock, *Organizational Science*. London : Tavistock Publications.
- Trochim , W. (1998). Donald T. Campbell and Research Design . *American Journal of Evaluation* , 407-409.
- Trochim , W., & Cook , J. (1992). Pattern Matching in Theory-Driven Evaluation: A Field Example from Psychiatric Rehabilitation. En H. Chen , & P. Rossi, *Research Strategies and Methods* (págs. 49-69). Westport, CT: Greenwood.
- Trochim, W. (2006 , October 20). *Web Center for Social Research Methods*. Retrieved from Positivism & Post-Positivism : <https://www.socialresearchmethods.net/kb/positvsm.php>
- Tyler , R. (1935). Evaluation: A challenge to progressive education. *Educational Research Bulletin* , 9-16.
- Tyler, R. W. (1942). General Statement on Evaluation . *Journal of Educational Research*, 492-501.

- Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. Chicago : University of Chicago press.
- Ulrich , W. (1981). A critique of pure cybernetic reason: the Chilean experience with cybernetics . *Journal of Applied Systems Analysis* , 33-59.
- Ulrich , W. (1993). Some Difficulties of Ecological Thinking, Considered from a Critical Systems Perspective: A Plea for Critical Holism. *Systems Practice*, 583-611.
- Ulrich , W. (1996). Critical Systems Thinking for Citizens. En R. Flood , & N. Romm, *Critical Systems Thinking: current research and practice* (págs. 165-179). New York: Plenum Press.
- Ulrich , W. (1996a). Critical Systems Thinking for Citizens. En R. Flood , & N. Romm, *Critical Systems Thinking: current research and practice* (págs. 165-179). New York: Plenum Press.
- Ulrich , W. (1996b). *A primer to Critical Systems Heuristics for Action Researchers*. Hull: The Centre for System Studies.
- Ulrich , W. (1998). Systems Thinking as if People Mattered: Critical Systems Thinking for Citizens and Managers. *Lincoln School of Management*.
- Ulrich , W. (2000). Reflective practice in the Civil Society: The Contribution of Critically Systemic Thinking. *Reflective Practice*, 247-268.
- Ulrich , W. (2003). Beyond methodology choice: critical systems thinking as critically systemic discourse. *Journal of the Operational Research Society* , 325-342.
- Ulrich , W. (2004). C. West Churchman, 1913-2004 (obituary). *Journal of the Operational Research Society* , 1123-1129.
- Ulrich , W. (2005a). A mini-primer of critical systems heuristics. Rev. version of "Critical systems heuristics,". En H. G. Daellenbach , & R. L. Flood, *The Informed Student Guide to Management Science* (pág. 72f). London: Thompson Learning.
- Ulrich , W. (2005b). A mini-primer of boundary critique. Rev. version of "Boundary critique,". En H. G. Daellenbach, *The Informed Student Guide to Management Science, London* (pág. 41f). London : Thomson Learning.
- Ulrich , W., & Reynolds, M. (2010). Critical systems heuristics. En M. Reynolds , & S. Holwell, *Systems Approaches to Managing Change: A Practical Guide* (págs. 243-292). London: Springer.
- Ulrich, W. (1983). *Critical Heuristics of Social Planning*. Chichester : John Wiley & Sons.
- Ulrich, W. (1983). *Critical Heuristics of Social Planning* . Berne : Paul Haupt .
- Ulrich, W. (1987). Critical Heuristics of Social Systems Design. *European Journal of Operational Research*.
- UNESCO . (2017). *UNESCO Learning to Live Together*. Obtenido de <http://www.unesco.org/new/en/social-and-human-sciences/themes/international-migration/glossary/poverty/>
- UNICEF. (2009). *The Hermes Programme for Conflict Management at School*. Panama: UNICEF.
- Unidad de Víctimas. (1 de December de 2017). *Red Nacional de Información* . Obtenido de <https://cifras.unidadvictimas.gov.co/Home/Desplazamiento>

- Von Bertalanffy, L. (1950). The theory of open systems in physics and biology. *Science*, 23-29.
- Von Bertalanffy, L. (1956). General Systems Theory. *General Systems*, Vol 1, 1-10.
- Walton, M. (2016). Expert views on applying complexity theory in evaluation: opportunities and barriers. *Evaluation*, Vol 22 (4), 410-423.
- Walton, M. (2016). Experts views on applying complexity theory in evaluation: opportunities and barriers. *Evaluation*, 410-423.
- Wandersman, A., & Snell-Johns, J. (2005). Empowerment evaluation: Clarity, dialogue, and growth. *American Journal of Evaluation*, 26, 421-428.
- Ward, A., Atkins, D. C., Lepper, M. R., & Ross, L. (2011). Affirming the Self to Promote Agreement With Another: Lowering a Psychological Barrier to Conflict Resolution. *Personality and Social Psychology Bulletin*, Vol 20, 1-13.
- Warfield, M. E. (1994). A cost-effectiveness analysis of early intervention services in Massachusetts: implications for policy. *Educational Evaluation and Policy Analysis*, 16, 87-99.
- Weiss, C. (1972). *Evaluation Research*. Englewood Cliffs: Prentice Hall.
- Weiss, C. (1987). Where Politics and Evaluation Research Meet. En D. Palumbo, *The Politics of Program Evaluation*. Newbury Park; CA: Sage Publications.
- Weiss, C. (1995). Nothing as Practical as Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives for Children and Families. En J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. Weiss, *New Approaches to Evaluation Community Initiatives: Volumen 1, Concepts, Methods, and Contexts*. Washington, DC: The Aspen Institute.
- Weiss, C. (1997). Theory-based Evaluation: past, present, and future. *New Directions for Evaluation*, 41-55.
- Werner, U., & Reynolds, M. (2010). Critical systems heuristics. En M. Reynolds, & S. Holwell, *Systems approaches to managing change: A practical guide*. London : Springer.
- Westhorp, G. (2012). Using complexity-consistent theory for evaluating complex systems. *Evaluation*, 405-420.
- Westhorp, G. (2012). Using complexity-consistent theory for evaluating complex systems. *Evaluation*, 405-420.
- Westhorp, G. (2013). Developing complexity-consistent theory in a realist investigation. *Evaluation*, 364-382.
- Wilke, A., & Mata, R. (2012). Cognitive Bias. En V. S. Ramachandran, *The Encyclopedia of Human Behavior* (págs. 531-535). Elsevier.
- Williams, B., & Hummelbrunner, R. (2011). *Systems Concepts in Action*. Stanford, California: Stanford University Press.
- Williams, B., & Hummelbrunner, R. (2011). *Systems Concepts in Action: A Practitioner's toolkit*. Stanford: Stanford University Press.
- Williams, B., & van't Hof, S. (2014). *Wicked Solutions. A Systems Approach to Complex Problems*. Gumround.

- Wilson , A. L., & Cervero , R. M. (1996). Paying attention to the people work when planning educational programs for adults . *New Directions for Adult and Continuing Education* , 69, 91-99.
- Wolf-Branigin , M. (2013). *Using Complexity Theory for Research and Program Evaluation* . Oxford : Oxford University Press .
- WordPress. (2013, June 11). *The Focusing-Effect: Why We End Up Making the Wrong Decisions*. Retrieved from Reflectd On the Mind: <http://reflectd.co/2013/06/11/the-focusing-effect-biased-decision-making/>
- Zhang , G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis , K. (2011). Using the Context, Input, Process, and Product Evaluation Model (CIPP) as a Comprehensive Framework to Guide the Planning, Implementation, and Assessment of Service-learning Programs. *Journal of Higher Education Outreach* , 15, 57-83.
- Zukoski, A. P., & Bosserman , C. (2017). Essentials of Participatory Evaluation. En D. Fetterman , L. Rodriguez-Campos, & A. P. Zukoski, *Collaborative, Participatory, and Empowerment Evaluation* (págs. 48-57). New York : Guilford Press.
- Zukoski, A. P., & Luluquisen , M. (2002). Participatory Evaluation: What is it? Why do it? What are the challenges? *Community-Based Public Health Policy and Practice*, 5 , 3-8.
- Zukoski, A. P., Bosserman , C., & Baechler, C. (2017). A Participatory evaluation of a Community Health Improvement Initiative. En D. Fetterman , L. Rodriguez-Campos, & A. P. Zukoski, *Collaborative, Participatory, and Empowerment Evaluation* (págs. 57-66). New York : Guilford Press.

Appendices

Appendix A – CSH Questionnaires (workshop activity 1)

CSH questionnaires were the first activity conducted with the stakeholders of the Hermes Program. Each one of the stakeholders (teachers, parents, students, former students of the program, the school's administrative staff, and the program consultants) who were considered for conducting the workshop. This appendix includes the questionnaires in Spanish, as this was the original language in which they were applied, and in English so that the reader might be able to read them.

Consultores – Consultants Questionnaire

Con el siguiente cuestionario deseamos conocer cuáles son sus percepciones sobre diversos aspectos del Programa Hermes. Por esta razón, usted contará con 50 minutos para desarrollar las preguntas que aquí se presentan. Todas las preguntas son abiertas y puede escribir tanto y tan detalladamente como lo desee. Sus respuestas son confidenciales y serán de gran ayuda para evaluar, enriquecer y mejorar el programa por lo que le pedimos responder con la mayor honestidad posible. *(With the following questionnaire we would like to know your perceptions about different issues related to the Hermes Program. For this reason, you will have 50 minutes to answer the questions presented in the following questionnaire. All the questions are open and you are allowed to write as much and as detailed as you wish. Your answers are confidential and will be very helpful for evaluating, enriching and improving the program, for this reason we ask you to answer as honest as possible).*

Muchas gracias por su participación. *(Thank you so much for your participation).*

1. ¿Cómo considera que es el compromiso de los docentes y coordinadores de las instituciones educativas que hacen parte del programa? *(How do you consider is the degree of commitment of the teachers and the administrative staff of the schools that are part of the program?)*
2. ¿Quién o quienes cree usted, se están viendo beneficiados por el programa? ¿Cree que estas son las personas que deberían ser las beneficiadas? o ¿Considera usted que los beneficiarios deberían ser otros? o ¿Algunas personas adicionales a las que ya se benefician? *(Who do you believe is being benefited by the program? Do you believe these individuals are the ones who should be benefited by it? Aside of the current beneficiaries do you believe someone else should be benefited by it?)*
3. ¿Qué factores considera usted son los más relevantes o clave para que la ejecución del programa tenga éxito? *(What factors do you consider are key for a successful execution of the program?)*
4. ¿Cuáles cree que son las características sociales, políticas, económicas y culturales que rodean al programa tanto a nivel de comunidad, localidad como a nivel

institucional? ¿Considera que alguna o varias tienen una influencia directa en el desempeño del programa tanto positiva como negativamente? (*Which are the social, political, economic, and cultural issues surrounding the program at a community, local and institutional level? Do you believe that some of these have a direct influence in the performance of the program?*)

5. ¿Desde su punto de vista, cuáles han sido algunas mejoras que se han presentado en las instituciones educativas que asiste? (*From your point of view, which have been some improvements that have taken place as a result of the program in the schools that you support?*)
6. ¿Cómo ha sido la participación por parte de los estudiantes? ¿Cómo se garantiza el compromiso y la motivación sostenida a lo largo de la participación de estos en el programa? (*How has been the student's participation in the program? How do you think a sustained motivation and commitment could be guaranteed throughout the participation in this type of programs?*)
7. Desde su posición como consultor, ¿Cómo considera que ha sido su labor en la implementación del programa? ¿Cuáles han sido los retos más grandes que ha tenido que enfrentar? (*From your position as a consultant, How do you think has been your work in the implementation of the program? Which have been the biggest challenges you have had to face so far?*)
8. Este espacio se encuentra disponible para comentarios adicionales que usted desee hacer (si lo desea también puede usar el respaldo de la hoja) (*This space is free for you to write additional comments you wish to make*).

Coordinadores y orientadores – Administrative Staff Questionnaire

Con el siguiente cuestionario deseamos conocer cuáles son sus percepciones sobre diversos aspectos del Programa Hermes. Por esta razón, usted contará con 50 minutos para desarrollar las preguntas que aquí se presentan. Todas las preguntas son abiertas y puede escribir tanto y tan detalladamente como lo desee. Sus respuestas son confidenciales y serán de gran ayuda para evaluar, enriquecer y mejorar el programa por lo que le pedimos responder con la mayor honestidad posible. *(With the following questionnaire we would like to know your perceptions about different issues related to the Hermes Program. For this reason, you will have 50 minutes to answer the questions presented in the following questionnaire. All the questions are open and you are allowed to write as much and as detailed as you wish. Your answers are confidential and will be very helpful for evaluating, enriching and improving the program, for this reason we ask you to answer as honest as possible).*

Muchas gracias por su participación. (Thank you so much for your participation).

1. ¿Qué lo motivó a que su institución fuera parte del Programa Hermes? y ¿Qué lo motiva a usted a ser parte del programa? *(What motivated you to involve your institution in the Hermes Program? What motivates you to be part of the program?).*
2. ¿Cuáles cree que son las características sociales, políticas, económicas y culturales que rodean al programa tanto a nivel de comunidad, localidad como a nivel institucional? ¿Considera que alguna o varias tienen una influencia directa en el desempeño del programa? *(In your opinion, Which are the social, political, economic, and cultural issues surrounding the program at a community, local and institutional level? Do you believe that some of these have a direct influence in the performance of the program?).*
3. Teniendo en cuenta las características sociales, económicas, políticas y culturales que rodean a su institución, ¿Considera que el enfoque del programa Hermes es el apropiado? ¿Cree debería trabajar aspectos adicionales o tener un enfoque diferente? De ser así, ¿cuáles? *(Taking into account the social, economic, politic and cultural*

factors surrounding your school, do you perceive the scope of the Hermes Program as being adequate? Do you think additional issues should be included, or that a different scope should be given to it? In case you give an affirmative answer to this question, could you please give more detail).

4. ¿Cuál cree que ha sido el impacto del programa en la institución o en la comunidad educativa? *(What do you believe has been the impact of the program at an institutional and community level?).*
5. ¿Cómo considera que es el compromiso con el programa Hermes por parte de los docentes y estudiantes de la institución donde usted se desempeña como coordinador u orientador? *(How do you consider is the degree of commitment of the teachers and students in your school?).*
6. ¿Cómo cree que los otros miembros de la comunidad educativa perciben el programa (aquellos que no pertenecen a él)? ¿Cómo cree que esas percepciones se articulan y cuál es la dominante entre ellas? En otras palabras, ¿Cómo cree que los otros actores perciben al programa y los objetivos del mismo? *(How do you believe other members of the community which are not part of the program perceive it? How do you believe those perceptions articulate and which is the dominant? In other words, how do you believe other actors perceive the program and its goals?).*
7. ¿Qué opinión tiene de la labor del personal de la Cámara de Comercio y del acompañamiento que realizan? *(What is your opinion of the work of the Chamber of Commerce personnel?).*
8. Desde su posición como coordinador u orientador, ¿Cómo considera que ha sido su labor en la implementación del programa? ¿Cuáles han sido los retos más grandes que ha tenido que enfrentar? *(From your position as being part of the administrative staff of the school, How do you think has been your role in the implementation of the program? Which have been the biggest challenges you have faced so far?).*
9. Este espacio se encuentra disponible para comentarios adicionales que usted desee hacer. *(This space is free for you to make additional comments you consider necessary).*

Docentes – Teacher's Questionnaire

Con el siguiente cuestionario deseamos conocer cuáles son sus percepciones sobre diversos aspectos del Programa Hermes. Por esta razón, usted contará con 50 minutos para desarrollar las preguntas que aquí se presentan. Todas las preguntas son abiertas y puede escribir tanto y tan detalladamente como lo desee. Sus respuestas son confidenciales y serán de gran ayuda para evaluar, enriquecer y mejorar el programa por lo que le pedimos responder con la mayor honestidad posible. *(With the following questionnaire we would like to know your perceptions about different issues related to the Hermes Program. For this reason, you will have 50 minutes to answer the questions presented in the following questionnaire. All the questions are open and you are allowed to write as much and as detailed as you wish. Your answers are confidential and will be very helpful for evaluating, enriching and improving the program, for this reason we ask you to answer as honest as possible).*

Muchas gracias por su participación. *(Thank you so much for your participation).*

1. ¿Qué lo motivó a ser parte del Programa Hermes? *(What motivated you to be part of the Hermes Program?)*.
2. ¿Considera que el enfoque del programa es el apropiado? ¿Cree que debería trabajar aspectos adicionales o tener un enfoque diferente? De ser así, por favor mencione cuales y por qué. *(Do you believe the scope of the program is adequate? Do you think it should include additional issues or have a different scope? In case of doing so, could you please describe which ones and why).*
3. ¿Cómo considera que es el compromiso con el programa Hermes por parte de los coordinadores, estudiantes y padres de familia de la institución donde usted se desempeña como docente? *(How do you consider is the degree of commitment of the students, administrative staff and parents in the school you work at?)*.
4. ¿Cree que el programa ha tenido éxito en su institución (comunidad, institución o salón de clase)? ¿Cuáles cree que son las variables que determinan el éxito del programa en su comunidad? *(Do you believe the program has been successful in the*

school you are working at (community, institution and classroom)? In your opinion which are the variables that determine the success of the program in your community?).

5. *¿Qué opinión tiene de la capacitación que la Cámara de Comercio le brinda a los docentes? Por favor justifique su respuesta. (What is your opinion about the training given to the teachers by the Chamber of Commerce?).*
6. *¿Cuáles cree que son las características sociales, políticas, económicas y culturales que rodean al programa tanto a nivel de comunidad, localidad como a nivel institucional? ¿Considera que alguna o varias tienen una influencia directa en el desempeño del programa? (Which are the social, political, economic, and cultural issues surrounding the program at a community, local and institutional level? Do you believe that some of these have a direct influence in the performance of the program?).*
7. *¿Cómo cree que los otros miembros de la comunidad educativa perciben las mejoras que el programa está produciendo o debería estar produciendo? ¿Cómo cree que esas percepciones se articulan y cuál es la dominante entre ellas? En otras palabras, ¿Cómo cree que los otros actores perciben al programa y los objetivos del mismo? (How do you think other members of the school's community perceive the improvements produced by the program or the improvements it should be producing? How do you think these perceptions articulate and which one is the dominant among them? In other words, how do you think other actors perceive the program and its goals?).*
8. *¿Desde su posición como docente, como considera que ha sido su labor en la implementación del programa? ¿Cuáles han sido los retos más grandes que ha tenido que enfrentar? (From your position as a teacher, how do you think has been your work in the implementation of the Hermes Program? Which have been the biggest challenges you have faced so far?).*
9. *Este espacio se encuentra disponible para comentarios adicionales que usted desee hacer. (This space is available for you to make additional comments).*

Egresados – Former students of the program Questionnaire

Con el siguiente cuestionario deseamos conocer cuáles son sus percepciones sobre diversos aspectos del Programa Hermes. Por esta razón, usted contará con 50 minutos para desarrollar las preguntas que aquí se presentan. Todas las preguntas son abiertas y puede escribir tanto y tan detalladamente como lo desee. Sus respuestas son confidenciales y serán de gran ayuda para evaluar, enriquecer y mejorar el programa por lo que le pedimos responder con la mayor honestidad posible. *(With the following questionnaire we would like to know your perceptions about different issues related to the Hermes Program. For this reason, you will have 50 minutes to answer the questions presented in the following questionnaire. All the questions are open and you are allowed to write as much and as detailed as you wish. Your answers are confidential and will be very helpful for evaluating, enriching and improving the program, for this reason we ask you to answer as honest as possible).*

Muchas gracias por su participación. (Thank you so much for your participation).

1. ¿Qué lo motivó a ser parte del Programa Hermes? *(What motivated you to be part of the Hermes Program?)*.
2. ¿Considera que el enfoque del programa es el apropiado? ¿Cree que debería trabajar aspectos adicionales o tener un enfoque diferente? De ser así, ¿cuáles? *(Do you think the scope of the program is appropriate? Do you believe it should work additional aspects or have a different scope? Which ones?)*.
3. ¿Cree que el programa ha tenido algún impacto en la comunidad? *(Do you think the program has had an impact on the community?)*.
4. Habiendo terminado su participación en el programa, ¿Considera que las habilidades aprendidas han sido de utilidad para su vida? ¿Cuáles son las habilidades que más destaca y por qué? *(Having finished your participation in the program, do you believe what you have learnt has been useful for your daily life? Which do you consider are the abilities you highlight the most from what you learnt and why?)*.

5. ¿Qué factores considera usted son los más relevantes o clave para que la ejecución del programa tenga éxito? (*What factors you consider are key for a successful execution of the program?*).
6. ¿Cuál es su opinión sobre la labor que juegan los docentes, coordinadores y los padres de familia en la ejecución del programa? (*What is your opinion regarding the role of the teachers, administrative staff and parents in the execution of the program?*).
7. ¿Cuáles cree que son las características sociales, políticas, económicas y culturales que rodean al programa tanto a nivel de comunidad, localidad como a nivel institucional? ¿Considera que alguna o varias tienen una influencia directa en el desempeño del programa tanto positiva como negativamente? (*Which are the social, political, economic, and cultural issues surrounding the program at a community, local and institutional level? Do you believe that some of these have a direct influence in the performance of the program?*).
8. Este espacio se encuentra disponible para comentarios adicionales que usted desee hacer (si lo desea también puede usar el respaldo de la hoja). (*This space is available for you to make additional comments*).

Estudiantes- Students Questionnaire

Con el siguiente cuestionario deseamos conocer cuáles son sus percepciones sobre diversos aspectos del Programa Hermes. Por esta razón, usted contará con 50 minutos para desarrollar las preguntas que aquí se presentan. Todas las preguntas son abiertas y puede escribir tanto y tan detalladamente como lo desee. Sus respuestas son confidenciales y serán de gran ayuda para evaluar, enriquecer y mejorar el programa por lo que le pedimos responder con la mayor honestidad posible. *(With the following questionnaire we would like to know your perceptions about different issues related to the Hermes Program. For this reason, you will have 50 minutes to answer the questions presented in the following questionnaire. All the questions are open and you are allowed to write as much and as detailed as you wish. Your answers are confidential and will be very helpful for evaluating, enriching and improving the program, for this reason we ask you to answer as honest as possible).*

Muchas gracias por su participación. (Thank you so much for your participation).

1. ¿Qué te motivó a ser parte del Programa Hermes? *(What motivated you to be part of the Hermes Program?)*.
2. ¿Crees que los profesores y los coordinadores de tu colegio están comprometidos con las actividades del programa? Es decir, ¿Brindan los espacios, el apoyo y los horarios suficientes para realizar las actividades de Hermes? *(Do you think the teachers and administrative staff of your school are committed with the activities of the program? Do they provide adequate time frames, spaces, and support to conduct the activities of the program?)*.
3. ¿Crees que los temas que aborda el programa son apropiados? ¿Te gustaría que incluyeran otros temas? De ser así, por favor menciona cuales y por qué. *(Do you believe the issues tackled by the program are appropriate? Would you like to include other issues aside of the ones already included? Please mention which ones and why).*

4. ¿Cómo crees que es visto el programa por otros estudiantes y profesores que no están involucrados en el programa? ¿Cómo te hace sentir lo que los demás dicen del programa? (*How do you think the program is perceived by other teachers and students that are not involved in it? How does their opinion makes you feel?*).
5. ¿Qué es lo que más te gusta y lo que menos te gusta del programa? (*What do you like the least and the most from the program?*).
6. ¿Crees que el programa ha funcionado en tu colegio? ¿por qué? (*Do you think the program has worked in your school? Why?*).
7. ¿Tienes conocimiento de la presencia de algún grupo en tu colegio que impida que tú o tus compañeros puedan participar libre o activamente en el programa Hermes? De ser así, ¿Cuáles son las actitudes o comportamientos de estas personas hacia ti o tus compañeros? (*Are you aware of any group in your school which hinders your participation or that of your classmates in the Hermes program? If so, which are their attitudes and behaviors towards you or your classmates?*).
8. ¿Desde tu posición como estudiante, como consideras que ha sido tu labor en el programa? ¿Cuáles han sido los retos más grandes que has tenido que enfrentar? (*From your position as a student, how do you think has been your participation in the program? Which have been the biggest challenges you have had to face so far?*).
9. Este espacio se encuentra disponible para comentarios adicionales que usted desee hacer. (*This space is for you to write additional comments in case of having any*).

Padres de familia – Parents Questionnaire

Con el siguiente cuestionario deseamos conocer cuáles son sus percepciones sobre diversos aspectos del Programa Hermes. Por esta razón, usted contará con 50 minutos para desarrollar las preguntas que aquí se presentan. Todas las preguntas son abiertas y puede escribir tanto y tan detalladamente como lo desee. Sus respuestas son confidenciales y serán de gran ayuda para evaluar, enriquecer y mejorar el programa por lo que le pedimos responder con la mayor honestidad posible. *(With the following questionnaire we would like to know your perceptions about different issues related to the Hermes Program. For this reason, you will have 50 minutes to answer the questions presented in the following questionnaire. All the questions are open and you are allowed to write as much and as detailed as you wish. Your answers are confidential and will be very helpful for evaluating, enriching and improving the program, for this reason we ask you to answer as honest as possible).*

Muchas gracias por su participación. (Thank you so much for your participation).

1. ¿Qué opina sobre el Programa Hermes? ¿Qué opinión tiene sobre la gestión del programa en la Institución donde estudia su hijo(a)? *(What do you think about the Hermes Program? What do you think about the management it has been given at the school your child attends?).*
2. ¿Considera que los temas abordados por el programa son apropiados? ¿Le gustaría que incluyeran otros temas? De ser así, por favor menciona cuales y por qué. *(Do you think the issues tackled by the program are appropriate? Would you like it to include other issues? Which ones and why?).*
3. ¿Cuáles cree que son las características sociales, políticas, económicas y culturales que rodean al programa tanto a nivel de comunidad, localidad como a nivel institucional? ¿Considera que alguna o varias influyen directamente en el desempeño del programa (tanto positiva como negativamente)? *(Which are the social, political, economic, and cultural issues surrounding the program at a community, local and*

institutional level? Do you believe that some of these have a direct influence in the performance of the program?).

4. *¿Cuál cree que ha sido el impacto del programa en la comunidad? (What do you think has been the impact of the program in the community?)*
5. *¿Quién o quienes cree usted, se están viendo beneficiados por el programa? ¿Cree que estas son las personas que deberían ser las beneficiadas? o ¿Considera usted que los beneficiarios deberían ser otros? o ¿Algunas personas adicionales a las que ya se benefician? (Who do you think is being benefited by the program? Do you think they should be the beneficiaries, or someone else should? Do you consider additional potential beneficiaries to be taken into account?)*
6. *¿Qué lo motivó a ser un padre o madre activamente involucrado en el Programa Hermes? (What motivated you to become an actively involved parent in the Hermes Program?)*
7. *¿Cuáles cree usted que son los elementos que permiten que el programa tenga éxito en su institución? (Which elements do you believe are the ones that allow a successful implementation of the program at the school your child attends?)*
8. *Este espacio se encuentra disponible para comentarios adicionales que usted desee hacer (si lo desea también puede usar el respaldo de la hoja). (This is space is for you to write additional comments if necessary).*

Appendix B – Stories of Change Format

Formato- Actividad Cambio Significativo

Título de la Historia (*Title of the story*):

Quién estuvo involucrado cuando ocurrió la historia de cambio (*Who was involved in the story?*):

Descripción de la situación (*Describe the situation*):

¿Por qué la historia es significativa? (*Why is the story considered as significant?*)

Lecciones aprendidas o recomendaciones para el programa (*Which lessons were learnt and which might be some recommendations for the program*)

Appendix C – Six Thinking Hats Format

Formato – Actividad Sombreros Para Pensar

Ronda: _____

Grupo: _____

Sombrero Verde (Propuesta de Cambio) (*Green Hat*):

Sombrero Negro (Juicios Negativos) (*Black Hat*):

Sombrero Amarillo (Juicios Positivos) (*Yellow Hat*):

Sombrero Rojo (emociones) (*Red Hat*):

Sombrero Blanco (Datos e Información Objetiva) (*White Hat*):

Sombrero Azul (Síntesis y Conclusiones) (*Blue Hat*):

Appendix D – Student's Survey (Original version)

Por medio de la siguiente encuesta queremos conocer tus puntos de vista con respecto a diversos aspectos del programa Hermes. Por esta razón te pedimos que respondas con una X a cada una de las preguntas y con la mayor sinceridad posible, ya que esto contribuye a la mejora del programa.

Institución Educativa: _____

Curso: _____

Edad: _____ años

Hombre: _____ **Mujer:** _____

1. ¿En qué fase del programa te encuentras?

No hago parte del programa _____ Comenzando (menos de 6 meses) _____

Primer año _____ Finalizando (a punto de certificarme) _____

Ya estoy certificado _____

2. ¿Con quién vives?

Solo con mis papas _____ Con mis papas y mis hermanos _____

Con mis papás, hermanos y otros miembros de la familia (abuelos, tíos o primos) _____

Solo con mis abuelos _____ Otro (Por favor especifica con quién) _____

3. ¿Cómo calificarías tu relación con las personas con las que vives?

Muy mala _____ Mala _____ Regular _____ Buena _____ Muy Buena _____

4. ¿Cómo crees que es la percepción general del Programa Hermes en tu colegio?

Muy mala _____ Mala _____ Regular _____ Buena _____ Muy Buena _____ No sé _____

5. ¿Cómo crees que las personas con las que vives perciben el programa?

Muy malo _____ Malo _____ Regular _____ Bueno _____ Muy Bueno _____ No lo conocen _____ No sé _____

6. Por favor marca con una X qué tan de acuerdo estás con cada una de las siguientes afirmaciones.

	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Completamente de acuerdo
La institución educativa brinda los <u>espacios</u> adecuados para realizar las actividades del					
La institución educativa brinda los <u>tiempos</u> adecuados para realizar las actividades del					
El <u>rector</u> y los <u>coordinadores</u> están altamente comprometidos con el desarrollo del programa					
Los docentes de la institución están altamente comprometidos con el desarrollo del programa					
Los talleres con los consultores de la Cámara de Comercio tienen la <u>duración</u> y la <u>frecuencia</u> adecuada					
Los <u>temas</u> que abordan los talleres del Programa Hermes son <u>apropiados</u>					
Los consultores de la Cámara de Comercio te tratan bien y demuestran conocer los temas del programa					

7. ¿Durante este año escolar con qué frecuencia has tenido los siguientes comportamientos hacia tus compañeros?

	Nunca	Alguna vez	A veces	Muchas veces	Siempre
Los insultas o los criticas					
Les haces una broma o te burlas de ellos					
Los golpeas o los maltratas físicamente					
Los ignoras o los rechazas					
Les amenazas o chantajeas					
Les robas o les dañas sus cosas					

8. ¿Durante este año escolar con qué frecuencia has sido víctima de los siguientes comportamientos por parte de tus compañeros?

	Nunca	Alguna vez	A veces	Muchas veces	Siempre
Te insultan o te critican					
Eres objeto de burlas o bromas					
Eres objeto de golpes o maltrato físico					
Te ignoran o te rechazan					
Recibes amenazas o chantajes					
Te roban o te rompen tus cosas					

9. ¿Qué haces al observar las siguientes situaciones entre tus compañeros?

	No hago nada, no les doy mi incumbencia	Manifiesto mi desacuerdo pero no hago parte del conflicto	Provoco diálogos entre ellos	Los enfrento y peleo con ellos	Ayudo al que me parece mejor	Le digo al profesor
Un compañero se burla de otro						
Un compañero insulta a otro						
Un grupo de estudiantes lo rechaza a un compañero						
Un compañero amenaza a otro						
Uno o varios compañeros agreden físicamente a otro						
Hay una pelea entre varios compañeros						
Un compañero le roba o le quita sus cosas a otro						

10. ¿Qué es lo que más te gusta del Programa Hermes? (Puedes seleccionar hasta 3 respuestas)

El compromiso del colegio con Hermes ____ El apoyo de los consultores ____ El compromiso de mis compañeros con Hermes ____ Los talleres y temas de Hermes ____ El reconocimiento por participar en el programa ____ El impacto de Hermes en el colegio ____ El impacto de Hermes en mi vida ____

11. ¿Qué es lo que más te disgusta del Programa Hermes? (Puedes seleccionar hasta 3 respuestas)

El compromiso del colegio con Hermes ____ El apoyo de los consultores ____ El compromiso de mis compañeros con Hermes ____ Los talleres y temas de Hermes ____ El reconocimiento por participar en el programa ____ El impacto de Hermes en el colegio ____ El impacto de Hermes en mi vida ____

12. ¿Crees que la implementación del Programa Hermes en tu colegio ha sido?

Pésimo ____ Muy malo ____ Malo ____ Regular ____ Bueno ____ Muy bueno ____ Excelente ____

Appendix E – Consultants' Survey (Original Version)

A través del siguiente cuestionario deseamos conocer sus percepciones, opiniones y sugerencias con respecto a diferentes aspectos relacionados con la implementación del Programa Hermes en las instituciones educativas que se encuentran a su cargo. Por esta razón, le pedimos que conteste cada una de las preguntas de la manera más honesta posible. A continuación encontrará 20 preguntas, abiertas y cerradas. A menos que la pregunta lo indique, todas las preguntas cerradas son de única respuesta. Muchas gracias por su colaboración.

1. Por favor indique su género :

Hombre____ Mujer____ Prefiero no decirlo____

2. Por favor indique su edad:

3. ¿Cuál es su nivel de estudios?

Bachillerato____ Técnico____ Universitario____ Posgrado____

4. ¿Cuál es el área de sus estudios profesionales?

Administración de Empresas____ Arquitectura y Diseño____ Artes y Humanidades____
Ciencias____ Ciencias Sociales____ Derecho____
Economía____ Educación____ Áreas de la salud____
Otra____

5. ¿Cuánto tiempo lleva desempeñándose como consultor del Programa Hermes?

Menos de 1 año____ Entre 1 y 2 años____ Entre 2 y 3 años____ Entre 3 y 4 años____ Entre 4 y 5 años____ Más de 5 años____ No sabe/ no responde____

6. Por favor indique el número de colegios que asiste como consultor del Programa Hermes.

7. Describa brevemente ¿Qué considera usted se hace al interior de las instituciones educativas para mejorar las competencias de comunicación entre los participantes del Programa Hermes?

8. En su opinión, ¿Cuál considera usted, es el grado de compromiso de los diferentes actores de la comunidad educativa con el Programa Hermes?

	Muy Bajo	Bajo	Medio	Alto	Muy Alto
Personal					
Administrativo					
Docentes					
Estudiantes					
Padres de Familia					
Consultores					

9. ¿Qué factores considera usted limitan el compromiso de cada uno de los miembros de la comunidad educativa? ¿Qué se podría hacer para potenciar el compromiso?

10. De acuerdo con su experiencia, ¿En qué medida han mejorado las relaciones entre los siguientes grupos de actores en las instituciones que usted asiste?

	Muy Bajo	Bajo	Medio	Alto	Muy Alto
Estudiantes-Estudiantes					
Estudiantes-Docentes					
Estudiantes-Personal Administrativo					
Estudiantes-Padres de Familia					
Estudiantes-Consultor					
Docentes-Personal Administrativo					
Docentes-Consultor					
Padres de Familia-Personal Administrativo					

11. En su opinión, ¿Cuáles son las variables más importantes para una ejecución exitosa del Programa Hermes? (Seleccione máximo 5)

Tiempos y espacios proporcionados por las instituciones para realizar las actividades____
 Tiempo disponible por parte de docentes y estudiantes____ Compromiso e interés de los estudiantes____ Compromiso e interés de los docentes____ Compromiso y apoyo del personal administrativo (directivos)____ Capacidades de los docentes____ Calidad y contenido de las capacitaciones para docentes____ Calidad, claridad, dinamismo y contenido del material que se trabaja con estudiantes____ Comunicación de horarios y actividades____ Sensibilización de la comunidad educativa____ Perfil del consultor____ Todas las anteriores____

12. ¿Considera usted que los requerimientos administrativos (metas, indicadores, papeleo, entre otros) de la Cámara de Comercio limitan la calidad y el impacto del Programa?

Si____ No____ No sabe/no responde____

13. Por favor justifique su respuesta para la pregunta anterior (Pregunta 12)

14. En su opinión, ¿Qué tan importante es el compromiso del personal de la Cámara de Comercio en el éxito del Programa Hermes?

	1	2	3	4	5	
No es muy importante						Muy importante

15. Por favor seleccione una respuesta para cada una de las siguientes preguntas:

	Si	No	Tal vez	No sabe/no responde
Si las condiciones apropiadas se dieran, ¿Contemplaría usted la posibilidad de seguir trabajando como consultor del programa para el año 2017?				
¿Ve usted su labor como consultor en el programa como parte de un proyecto de vida en el largo plazo?				
¿Considera usted que una alta rotación en los consultores afecta negativamente el desempeño del programa?				
¿Cree usted que el programa tendría un mayor impacto si tuviera una mayor dedicación horario en los colegios y disminuyera el número de colegios vinculados?				

16. ¿Qué actividades cree usted motivarían a los padres a involucrarse más activamente en el Programa Hermes? (Por favor seleccione máximo 2)

Talleres con temas de liderazgo y emprendimiento para padres___ Actividades de convivencia y comunicación para padres e hijos___ Envío de reportes periódicos sobre el desempeño de sus hijos en el programa___ El reconocimiento público de la labor que llevan a cabo sus hijos___ Desarrollo de actividades con otras miembros de la comunidad que rodean a las instituciones educativas___ Trabajo conjunto con otras instituciones educativas en el Programa Hermes___ Ninguna de las anteriores___ Todas las anteriores___

17. ¿Se encuentra usted satisfecho con la estructura actual del programa?

Si___ No___ Tal vez___ No sabe/No responde___

18. Por favor justifique su respuesta a la pregunta anterior (Pregunta 17)

19. ¿Cómo percibe usted la calidad general del Programa Hermes al momento de compararlo con otros programas enfocados en la resolución de conflictos escolares?

	1	2	3	4	5	
Extremadamente Baja						Excelente

20. Este espacio se encuentra disponible para comentarios adicionales que desee hacer sobre el programa.

Appendix F – Student’s survey correlation analysis

	Colegio	Grado	Edad	Genero	Relaci~m	Perce~ol	Percep~m	Espacios	Tiempos	Compro~s	Compro~e
Colegio	1										
Grado	0.1041	1									
Edad	0.0862	0.728	1								
Genero	-0.0498	0.088	-0.0325	1							
RelacionFam	-0.0514	-0.0153	-0.0749	-0.0195	1						
Percepcio~o	-0.0311	-0.0699	-0.036	0.0179	0.0594	1					
Percepcion~	-0.0069	0.0719	0.0282	0.0249	-0.0642	-0.1494	1				
Espacios	-0.2636	-0.1163	-0.1472	-0.0048	0.1377	0.1858	-0.0267	1			
Tiempos	-0.2065	-0.1266	-0.1345	-0.0549	0.1243	0.1539	-0.0288	0.6123	1		
Compromiso	-0.2211	-0.1325	-0.1322	0.0079	0.0723	0.1461	-0.0589	0.3977	0.4595	1	
Compromiso	-0.1223	-0.1629	-0.1298	-0.067	0.1398	0.1658	-0.0253	0.357	0.3431	0.4265	1
TiemposTal~	-0.1812	-0.1155	-0.1567	-0.0246	0.1687	0.1591	-0.0333	0.3123	0.4532	0.2958	0.29
TemasTalle~	-0.0983	-0.1185	-0.1641	-0.0435	0.0694	0.1463	-0.0538	0.2909	0.36	0.373	0.3629
TratoConsu~	-0.1107	-0.111	-0.1568	-0.0247	0.0298	0.2053	-0.067	0.2623	0.2827	0.2982	0.3392
InsultosRec	0.0155	0.2247	0.2417	0.0036	0.0796	0.0381	-0.0476	0.0371	0.0929	0.1004	-0.0251
BurlasRec	-0.0205	0.1462	0.1685	0.0618	0.1112	-0.002	-0.038	0.0679	0.0997	0.1176	-0.0052
GolpesRec	-0.1412	0.1722	0.1747	0.1223	0.1066	-0.0312	-0.0382	0.1071	0.1313	0.1923	0.0892
RechazoRec	-0.0814	0.1819	0.1631	0.0116	0.1	-0.0169	-0.0417	0.1392	0.1926	0.1572	0.0152
AmenazasRe	-0.0405	0.212	0.2023	0.0248	0.07	-0.0195	-0.0502	0.03	0.0562	0.0997	0.0423
RobosRec	0.0319	0.2446	0.2126	-0.0122	0.051	0.0259	-0.0083	0.0496	0.0684	0.0639	-0.0054
Obs_burlas	-0.081	-0.0377	-0.0281	0.0635	0.0705	-0.0174	-0.0157	0.0936	0.0154	-0.0414	-0.0291
Obs_insultos	-0.0735	-0.0393	-0.0421	0.0792	0.0408	0.0541	-0.0237	0.0605	0.0122	-0.048	-0.0616
Obs_rechazo	0.0091	0.0114	0.0122	0.0587	0.0615	0.0282	-0.0003	0.0312	-0.0713	-0.0486	-0.0218
Obs_amenaz	-0.0644	0.0521	-0.0212	0.1647	0.0651	0.0564	0.0348	0.0546	0.0225	0.0056	0.0103
Obs_golpes	-0.07	0.0676	0.059	0.1119	0.0735	0.0363	0.0077	0.0805	0.0189	-0.0249	-0.0026
Obs_peleas	-0.0779	0.0214	-0.0336	0.088	0.0831	0.0468	0.0188	0.0775	0.0043	0.0107	0.0293
Obs_robos	-0.0797	0.0347	-0.0184	0.0246	0.0574	0.055	-0.0082	0.061	0.0092	0.0249	0.0366
Percepcio~al	0.0345	-0.0232	-0.0646	-0.0664	-0.03	-0.0228	-0.0136	0.0249	0.0327	0.0629	0.0483
Vivienda	-0.0397	0.0239	0.0255	0.0178	0.0318	-0.0207	0.0215	0.0036	0.0605	0.0393	-0.0177
Consultor	0.9743	0.1029	0.0574	-0.0625	-0.0503	-0.0146	0.0122	-0.2672	-0.1886	-0.2253	-0.1055
Fase	0.4238	-0.0026	-0.0238	-0.0858	-0.1069	0.0574	0.0046	-0.0819	-0.0486	-0.0667	-0.003
Disgustos	-0.0306	-0.0866	-0.0569	0.0006	0.0305	-0.0046	-0.0325	0.0167	-0.0002	0.0265	0.0349
Gustos	0.09	0.1281	0.1131	0.012	-0.0474	-0.026	-0.0224	-0.0634	-0.0555	-0.0903	-0.1195
InsultosOt	-0.0695	-0.1788	-0.1688	-0.0456	-0.099	-0.1284	0.0496	-0.0539	-0.0504	-0.0978	-0.0188
BurlasOt	0.0292	-0.0275	-0.0099	-0.1795	-0.0564	-0.1135	0.0427	-0.0615	-0.0315	-0.0335	-0.0548
GolpesOt	0.0436	-0.1853	-0.1235	-0.1856	-0.0619	-0.0101	0.0164	-0.0787	-0.0264	-0.0931	-0.0993
RechazoOt	-0.066	-0.1326	-0.1009	0.0163	-0.0474	-0.0936	-0.0124	-0.0864	-0.0977	-0.0465	-0.0527
AmenazaOt	0.0223	-0.1287	-0.1155	-0.0697	-0.109	-0.06	0.0465	-0.0599	-0.0413	-0.0534	-0.0513
RobosOt	0.0533	-0.1022	-0.0884	-0.0744	-0.0849	-0.0017	-0.0581	-0.0728	-0.0594	-0.0733	-0.028

	Tiemp~es	TemasT~s	TratoC~s	Insult~c	Burlas~c	Golpes~c	Rechaz~c	Amenaz~c	RobosRec
Colegio									
Grado									
Edad									
Genero									
RelacionFam									
Percepcio~ol									
Percepcion~m									
Espacios									
Tiempos									
Compromiso~s									
Compromiso~e									
TiemposTal~	1								
TemasTalle~	0.4689	1							
TratoConsu~	0.4933	0.6164	1						
InsultosRec	0.0026	-0.0689	-0.0569	1					
BurlasRec	-0.0034	-0.0482	-0.0813	0.5982	1				
GolpesRec	0.021	0.0307	0.018	0.4581	0.5809	1			
RechazoRec	0.0534	0.0545	0.0413	0.5447	0.5325	0.5445	1		
AmenazasRe	-0.0173	-0.0153	-0.0126	0.3901	0.4471	0.6434	0.5433	1	
RobosRec	-0.0056	-0.0228	-0.0052	0.4129	0.3586	0.4641	0.4329	0.5409	1
Obs_burlas	0.0133	-0.0245	-0.0009	-0.0655	-0.0629	-0.0877	-0.0814	-0.1254	-0.105
Obs_insultos	-0.0096	-0.0299	-0.0206	-0.057	-0.051	-0.0958	-0.0963	-0.0767	-0.0866
Obs_rechazo	0.0084	-0.0429	0.01	-0.0688	-0.0925	-0.1024	-0.0629	-0.1083	-0.0984
Obs_amenaz	0.0122	0.0133	0.115	-0.058	-0.0601	-0.0273	-0.0264	-0.0712	0.0189
Obs_golpes	0.022	0.0193	0.0803	-0.1092	-0.0536	-0.0291	-0.0494	-0.057	-0.0777
Obs_peleas	0.0468	0.0012	0.0483	-0.0556	-0.0482	-0.0735	-0.1154	-0.0895	-0.0143
Obs_robos	0.0609	-0.0059	0.089	-0.0591	-0.0873	-0.0827	-0.0968	-0.0915	-0.0953
Percepcio~a	-0.0061	0.0573	-0.0237	-0.0267	0.0319	0.0124	0.0126	0.027	-0.0016
Vivienda	-0.002	0.0374	0.0218	0.0197	0.0357	0.0001	0.0248	-0.0056	0.0187
Consultor	-0.163	-0.0789	-0.0974	-0.0055	-0.0402	-0.1823	-0.117	-0.0826	0.0194
Fase	-0.0196	0.0505	0.0721	-0.0642	-0.0351	-0.1391	-0.0714	-0.0217	0.0227
Disgustos	-0.0183	-0.0098	0.0425	-0.0102	0.0532	0.0061	0.0218	-0.0135	-0.0399
Gustos	-0.0226	-0.0576	-0.0222	0.0037	0.0324	-0.007	-0.0042	-0.0143	0.0027
InsultosOt	-0.0299	-0.0262	0.0276	-0.3922	-0.2535	-0.1643	-0.1875	-0.1662	-0.1927
BurlasOt	-0.0184	0.038	0.0573	-0.2635	-0.2256	-0.1599	-0.1452	-0.1694	-0.1949
GolpesOt	0.0197	0.0167	0.0389	-0.2526	-0.2519	-0.4602	-0.2852	-0.42	-0.2755
RechazoOt	-0.0185	-0.0353	-0.0446	-0.3226	-0.2557	-0.2214	-0.4009	-0.3221	-0.2928
AmenazaOt	-0.0323	-0.0343	-0.0537	-0.2201	-0.1929	-0.3987	-0.3075	-0.4922	-0.3515
RobosOt	-0.0123	-0.0962	-0.0476	-0.2532	-0.2283	-0.3869	-0.3122	-0.418	-0.3734

	Obs_bu~s	Obs_in~s	Obs_re~o	Obs_am~s	Obs_go~s	Obs_pe~s	Obs_ro~s	Perce~al	Vivienda	Consul~r	Fase
Colegio											
Grado											
Edad											
Genero											
RelacionFam											
Percepcio~ol											
Percepcion~m											
Espacios											
Tiempos											
Compromiso~s											
Compromiso~e											
TiemposTal~s											
TemasTalle~s											
TratoConsu~s											
InsultosRec											
BurlasRec											
GolpesRec											
RechazoRec											
AmenazasRec											
RobosRec											
Obs_burlas	1										
Obs_insultos	0.6076	1									
Obs_rechazo	0.4379	0.4362	1								
Obs_amenaz	0.3706	0.4028	0.4072	1							
Obs_golpes	0.3605	0.3651	0.3444	0.5999	1						
Obs_peleas	0.3325	0.3773	0.3787	0.5019	0.5782	1					
Obs_robos	0.3229	0.3343	0.4143	0.4626	0.4653	0.541	1				
Percepcio~a	-0.036	-0.0397	-0.071	-0.0128	0.0074	-0.0346	-0.0168	1			
Vivienda	-0.0678	-0.0464	-0.0319	0.0224	-0.0318	0.0196	0.0213	0.0365	1		
Consultor	-0.0888	-0.0774	0.0027	-0.0615	-0.0742	-0.0568	-0.078	0.0349	-0.0282	1	
Fase	-0.0598	-0.0483	-0.1114	0.0117	-0.0142	0.0084	-0.03	0.1671	-0.0093	0.4463	1
Disgustos	-0.0411	-0.0405	-0.0483	-0.0468	-0.0105	-0.0185	-0.0279	-0.0187	0.0319	-0.0614	-0.0642
Gustos	0.0155	-0.0142	0.053	0.0619	0.0417	0.0404	-0.0073	-0.0423	-0.0422	0.084	-0.0289
InsultosOt	-0.0167	-0.0835	-0.0046	-0.0067	-0.0173	-0.0266	-0.0258	0.016	-0.0079	-0.0503	-0.0498
BurlasOt	-0.0587	-0.097	-0.0379	-0.0075	-0.0302	-0.0625	-0.0076	-0.0022	0.0212	0.0552	-0.0102
GolpesOt	0.0107	-0.0274	-0.0062	-0.0454	-0.0433	0.0069	0.0425	-0.0185	-0.0004	0.0868	0.0783
RechazoOt	0.0207	0.0275	0.0021	-0.0674	-0.0515	-0.0123	-0.0146	-0.0375	-0.0431	-0.0393	-0.0348
AmenazaOt	0.0154	0.0001	0.0534	-0.0828	-0.0366	-0.0093	-0.03	-0.0062	-0.0206	0.0525	0.0592
RobosOt	0.029	0.0094	0.04	-0.0607	0.003	-0.0042	-0.0192	0.0137	-0.0228	0.0796	0.0433

	Disgus~s	Gustos	Insult~t	BurlasOt	GolpesOt	Rechaz~t	Amenaz~t	RobosOt
Colegio								
Grado								
Edad								
Genero								
RelacionFam								
Percepcio~ol								
Percepcion~m								
Espacios								
Tiempos								
Compromiso~s								
Compromiso~e								
TiemposTal~s								
TemasTalle~s								
TratoConsu~s								
InsultosRec								
BurlasRec								
GolpesRec								
RechazoRec								
AmenazasRec								
RobosRec								
Obs_burlas								
Obs_insultos								
Obs_rechazo								
Obs_amenazas								
Obs_golpes								
Obs_peleas								
Obs_robos								
Percepcio~al								
Vivienda								
Consultor								
Fase								
Disgustos	1							
Gustos	-0.0201	1						
InsultosOt	-0.0584	0.0266	1					
BurlasOt	-0.0752	0.0255	0.495	1				
GolpesOt	-0.0479	0.0125	0.3359	0.424	1			
RechazoOt	0.0084	-0.0173	0.3361	0.2419	0.2813	1		
AmenazaOt	-0.0342	-0.0399	0.2793	0.31	0.4457	0.3909	1	
RobosOt	-0.0207	0.0139	0.2278	0.2506	0.3853	0.2739	0.5678	1

Appendix G – Consultants Survey Correlation Analysis

	genero	edad	Nivel_~s	Area_Est	Tiempo~s	Num_col	Comp_Adm	Comp_Doc	Comp_Est	Comp_P~s
genero	1									
edad	0.038	1								
Nivel_Estu~s	-0.2296	0.4034	1							
Area_Est	0.1711	-0.0149	-0.2794	1						
Tiempo_Con	-0.2875	0.43	0.4696	-0.0845	1					
Num_col						1				
Comp_Adm	0.0126	0.0256	-0.1744	0.3354	-0.0874		1			
Comp_Doc	0.0625	0.2336	0.0893	0.2139	0.0995		0.4209	1		
Comp_Est	-0.0563	0.0555	0.0153	-0.061	-0.0979		0.1395	0.5063	1	
Comp_Papas	0.3278	0.2305	-0.0698	0.2512	-0.1862		0.2091	0.0784	0.1593	1
Comp_Cons	0.2845	0.2247	0	0.0617	-0.2098		0.1716	0.4742	0.6046	0.1243
Rel_EstEst	0.1572	-0.0289	0.11	0	0.0318		0.2347	0.2694	0.2964	0.3379
Rel_EstDoc	0.0461	0.235	0.0188	0.2011	0.1311		0.2177	0.4493	0.3473	0.2876
Rel_EstAdm	0.1539	0.2512	0.1257	0.187	-0.0885		0.4487	0.2694	0.3291	0.3993
Rel_EstPad	0.0579	0.062	-0.0946	-0.1024	-0.2806		0.1668	-0.0869	0.1433	0.2652
Rel_EstCons	0.1833	-0.0996	-0.2994	-0.0392	-0.0798		0.0369	0.0688	0.3335	0.1123
Rel_DocAdm	0.1628	0.1006	-0.0443	0.1548	-0.2529		0.56	0.2985	0.1384	0.1011
Rel_DocCons	0.2372	0.0129	-0.1639	-0.2217	-0.1792		0.231	0.0547	0.0807	0.0426
Rel_PadCons	0.2754	0.2305	0.1124	0.2297	-0.003		0.3321	-0.0688	-0.08	0.3748
Vars_exito	0.3226	-0.0546	-0.1722	-0.1274	-0.3059		-0.0848	0.0744	0.0819	-0.1188
Req_Adm	0.3203	-0.0344	-0.4413	0.1644	-0.1983		-0.0161	-0.1301	-0.2282	0.2647
Imp_CompC	0.3842	0.0902	-0.3546	0.0343	-0.2897		0.3626	0.1754	-0.0952	0.1485
trab_2017	0.3203	0.1339	-0.1961	0.0685	-0.1157		-0.0161	0.1201	0.2822	0.2419
LargoPlazo	0.2845	0.3129	-0.2323	0.1655	0.014		0.0191	0.0474	-0.1351	0.2108
Rot	0.0239	0.2482	-0.039	-0.1756	-0.057		0.1007	-0.0358	-0.102	-0.1129
Hor_Col	-0.0385	0.0903	-0.1468	0.2417	-0.1901		0.1084	-0.199	-0.0135	0.3909
Mot_Pad	-0.2196	-0.0299	-0.061	-0.1567	-0.0645		-0.1052	-0.0327	0.0715	0.0309
Sat_Est	0.1655	0.254	-0.086	0.103	-0.0781		0.1996	0.1655	-0.1264	0.1304
Perc_Cal	0.3363	0.273	-0.0153	-0.0831	0.0452		0.2554	0.1495	-0.0645	0.0916

	Comp_C~s	Rel_Es~t	Rel_Es~c	Rel_Es~m	Rel_Es~d	Rel_Es~s	Rel_Do~m	Rel_Do~s	Rel_Pa~s	Vars_e~o	Req_Adm
genero											
edad											
Nivel_Estu~s											
Area_Est											
Tiempo_Cons											
Num_col											
Comp_Adm											
Comp_Doc											
Comp_Est											
Comp_Papas											
Comp_Cons	1										
Rel_EstEst	0.2044	1									
Rel_EstDoc	0.3059	0.5296	1								
Rel_EstAdm	0.2744	0.3871	0.3334	1							
Rel_EstPad	0.0659	0.2496	0.265	0.2674	1						
Rel_EstCons	0.2608	0.461	0.1985	0.0988	0.046	1					
Rel_DocAdm	0.1784	0.3119	0.2701	0.5903	0.1648	0.1791	1				
Rel_DocCons	0.2284	0.367	0.2505	0.073	0.2691	0.5787	0.367	1			
Rel_PadCons	-0.0131	0.1978	0.1237	0.6676	0.2684	-0.1073	0.3437	0.103	1		
Vars_exito	0	0.0089	0.0549	0.0764	-0.0268	0.0364	0.3016	0.0109	-0.0137	1	
Req_Adm	-0.243	-0.1294	-0.059	-0.1479	0.1113	0.0587	-0.0927	-0.0117	-0.0772	0.2026	1
Imp_CompC	0.3295	0.096	0.1909	0.1029	0.2528	0.147	0.3964	0.4827	0.1288	0.0431	0.3638
trab_2017	0.1519	-0.0575	0.1254	0.0246	0.0185	0.1321	0.0811	0.0175	0.1433	0.143	0.1026
LargoPlazo	0.0791	-0.2725	-0.0437	0.1343	-0.0513	0.0522	0.0686	-0.1038	0.1436	0.1506	0.1822
Rot	-0.0272	-0.3257	-0.0154	-0.0514	0.1971	-0.2581	0.0104	0.047	-0.0099	0.0047	0.2751
Hor_Col	0.1072	0.0185	0.0544	0.1819	0.113	0.146	0.0483	-0.0806	-0.0106	-0.0204	0.3372
Mot_Pad	-0.1347	-0.0571	-0.0844	0.0173	-0.0014	0.024	-0.1055	-0.0423	-0.0926	-0.0436	0.1167
Sat_Est	0.3197	0.227	0.122	0.0741	0.0837	0.0993	0.1481	0.2987	0.2321	-0.1464	0.1253
Perc_Cal	0.2126	0.1745	-0.0448	0.2186	0.0375	0.2672	0.2542	0.4718	0.2213	-0.0816	0.0479

	Imp_Co~C	tra~2017	LargoP~o	Rot	Hor_Col	Mot_Pad	Sat_Est	Perc_Cal
genero								
edad								
Nivel_Estu~s								
Area_Est								
Tiempo_Cons								
Num_col								
Comp_Adm								
Comp_Doc								
Comp_Est								
Comp_Papas								
Comp_Cons								
Rel_EstEst								
Rel_EstDoc								
Rel_EstAdm								
Rel_EstPad								
Rel_EstCons								
Rel_DocAdm								
Rel_DocCons								
Rel_PadCons								
Vars_exito								
Req_Adm								
Imp_CompC	1							
trab_2017	0.1498	1						
LargoPlazo	0.2788	0.3948	1					
Rot	0.338	-0.0688	-0.0996	1				
Hor_Col	0.1236	-0.1563	0.2241	0.1201	1			
Mot_Pad	-0.2572	0.1017	-0.1488	0.0767	0.0538	1		
Sat_Est	0.5148	0.0771	0.0457	0.1494	-0.0557	-0.0922	1	
Perc_Cal	0.3196	0.1077	0.3827	-0.1391	0.0038	0.0056	0.3059	1