DISSERTATION

USING CONCEPT MAPPING AS A TOOL FOR PROGRAM THEORY DEVELOPMENT

Submitted by

Rebecca Orsi

School of Education

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2011

Doctoral Committee:

Advisor: R. Brian Cobb

Jennifer A. Hoeting
Susan A. Lynham
Paul W. Speer

ABSTRACT

USING CONCEPT MAPPING AS A TOOL FOR PROGRAM THEORY DEVELOPMENT

The purpose of this methodological study is to explore how well a process called *concept mapping* (Trochim, 1989) can articulate the theory which underlies a social program. Articulation of a program's theory is a key step in completing a sound theory-based evaluation (Weiss, 1997a). In this study, concept mapping is used to articulate the outcomes domain of a program theory, using Chen's (1990) six domains for program theory as an organizing framework. A grassroots community organization in Denver, Colorado, provides context for the study. With reference to Dubin's (1978) distinctions for theoretical units as a guide, the results of concept mapping are analyzed to determine whether they are useful in building a program theory. Results are also are evaluated to determine whether they present a comprehensive, parsimonious (Whetten, 1989) and valid representation of outcomes from the community organizing intervention. Methodological and statistical considerations for using concept mapping are mentioned. The study concludes that concept mapping is a promising tool for theory articulation. Study limitations and opportunities for further research are also discussed.

ACKNOWLEDGEMENTS

I had a "village" of supporters as I conducted this study, and I would like to acknowledge them here. Sandy Brown and Mike Kromrey believe that academic research can benefit community organizations; this was an inspiration to me in my graduate school journey. Ryan Elmore provided assistance as I thought about how to manipulate data for MDS in R and it was he who reminded me of the "sparsity" challenge. Sarah Beutel, Brenda Ray, Jackie Yllescas, Kristee Paschall and Jon Stalls at Metro Organizations for People all provided logistical support for the study at MOP. I thank Jackie, especially, for her help in moving the project along. Stephanie Gut at the PICO National Network provided a very helpful document outlining the community organizing process used at MOP. Laura Sample McMeeking and Marc Winokur both provided model dissertations which helped me organize and format my own. Marc also generously provided the voice recorder I used during the interpretation session. I would like to acknowledge the "Tuesday Research Seminar", facilitated by Michael de Miranda, whose participants have discussed many iterations of this project over the last two years. Thanks also to Karen Atler, my "cohort of one", who has been a great source of moral support on the doctoral journey. I thank Dick and Dolores Orsi, my in-laws, who helped model for me the substantial benefits of an academic career. Many thanks to my parents, Jim and Nadine Hunt, who have always provided unfailing support to my career endeavors: academic, financial, emotional and lots of child care. And thanks again to

my mom for her help with the table of contents! I thank Jennifer Hoeting, who already sat on one graduate committee for me and was willing to sit on another. Sue Lynham provided many close readings of and detailed comments for drafts of this study. Thanks to Paul Speer for traveling so far to be a part of this work. Finally, I acknowledge Brian Cobb, who is the wisest and savviest PhD advisor any student ever had. I owe Brian additional thanks for allowing me to find my own way through this study and for not *really* retiring until I was done. My children, Renata and Carlos, popped sorting cards from perforated pages to help me assemble packets for the concept mapping process. Renata and Carlos were also the people who, more than anyone else, assumed without question that Mom's dissertation would get done! And, finally, to my husband Jared: thanks to you for listening without fail and for never doubting that this PhD is good for me, for our family and for the larger community.

For the leaders at Metro Organizations for People and at all grassroots community organizations: you do the hard work to change the world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 – Introduction

No religious festival or major holiday occurred that particular day, but on Thursday evening, November 10th, 2005, over 500 people filled Nuestra Señora de la Paz (Our Lady of Peace) Catholic Church in northeast Greeley, Colorado. They were attending a public meeting with Weld County District Attorney Ken Buck and Greeley Mayor Tom Selders. The organizing committee that orchestrated this event was composed of parishioners at the church; they were not professional event planners. They had only ten days to plan the meeting. How did they motivate so many people to attend? How did they get important politicians like the district attorney and the mayor to take an evening out of their busy schedules? And did this large meeting make a difference to individuals or to the civic community in any way?

…..

The brief anecdote about the public meeting in Greeley raises a salient question: how does one evaluate a complex social intervention? Theory-based evaluation (Chen, 1990; Weiss, 1972, 1998) may offer an answer. To be executed well, however, theory-based evaluation (TBE) requires the articulation of a robust theory that describes and explains the intervention in question. Using grassroots community organizing as an example of such a complex social intervention, this study explores the utility of concept mapping (Trochim, 1989) as a statistical and analytic tool for program theory development. Well-developed theory can in turn support strong theory-based evaluations.

Research Problem

On one hand, social interventions and programs are clearly present in the United States today. An example of one particularly complex intervention is the phenomenon of grassroots community organizing which involves itself in urban social and political processes in dynamic ways (Warren, 2001; Wood, 2002). Furthermore, there exists a growing consensus in the evaluation literature that theory should play a role in the evaluation of social programs (Chen, 1990; Weiss, 1998). On the other hand, there is also consensus in the literature that current theory-based evaluations could be better-executed (Rogers & Weiss, 2007; Weiss, 1997a). Despite a history spanning almost 40 years (Weiss, 1997b), theory-based evaluation continues to suffer from numerous challenges to doing it well (Weiss, 1997a). Three significant challenges include: (a) lack of clarity on the difference between black-box process-outcome evaluation and theory-based evaluation, (b) difficulty in constructing program theory, and (c) large time and resource requirements for TBE. Finally, Chen (1990) has suggested that statistical tools might be used in the development of program theory, particularly in the area of outcomes evaluation. Therefore, the purpose of this study is to explore the utility of concept mapping (Trochim, 1989) as a statistical and analytic tool for program theory development. More specifically, the study will use concept mapping to specify an outcomes domain in the context of Chen's (1990) six-domain framework for program theory. It will also examine whether or not the resulting outcomes domain meets several of Dubin's (1978) criteria for specifying the building blocks (i.e. units) of a theory. The study's explorations will contribute to current discussions in the literature about how theory-based evaluation might be improved via better theory articulation.

Research Questions

To achieve the purposes of the study mentioned above, the following research questions are posed.

1. How well does concept mapping assist in developing the outcomes domain of a program theory for a complex social intervention?

This first question will be answered by considering the following sub-questions which use Dubin's (1978, p. 37) "distinctions" for theoretical "units" as a guide.

   a. Does concept mapping produce potential units for incorporation into theory which (a) describe properties of objects instead of objects themselves and which (b) do not describe one-time events?
   b. Does concept mapping produce potential units which meet Dubin's four sets of mutually exclusive distinctions? That is, each unit must be classifiable as: (a) attribute or variable, (b) real or nominal, (c) primitive or sophisticated and (d) collective or member.

2. Are the potential units articulated by concept mapping collectively both parsimonious and comprehensive (Whetten, 1989) in describing the outcomes domain?
3. What evidence exists for the validity of programmatic outcomes as articulated by the concept mapping process?

Conceptual Framework

As suggested above, this study will investigate how researchers can use concept mapping within Chen's (1990) framework for articulating a program theory. Chen discusses six domains for a complete program theory. These domains are: treatment,

outcome, impact, intervening mechanism(s), implementation environment, and generalization. These domains are consistent with broader notions of what elements are required for a complete theory. For example, Patterson (1986) notes that a theory includes the following characteristics: (a) stated postulates and assumptions, (b) definitions of terms and concepts included in the theory, (c) statements of relationship among the terms and concepts in the theory, and, finally, (d) hypotheses/predictions that follow from the theory. Chen and Patterson align in this manner: Chen's domains of treatment, outcome and implementation environment are, in Patterson's language, "terms" (1986, p. xix) which require definition. Relationships between treatment and outcome are expressed in Chen's intervening mechanism domain. Expected impacts of treatments on outcomes (per Chen) constitute Patterson's predictions of the theory. And, finally, Patterson's assumptions must be made so the theory can describe to what contexts it might generalize.

The proposed study will focus on using concept mapping to build only the portion of a program theory that describes *outcomes*. Chen writes that "…normative outcome evaluation involves systematically identifying or clarifying a set of program goals or outcomes…" (1990, p. 54). Figure 1 below illustrates how Chen's program theory domains fit together and how concept mapping may work to articulate the outcomes domain as one element of a more comprehensive program theory. The *treatment* domain represents the intervention undertaken by a social program. *Outcomes* result from this intervention. An *intervening mechanism* is the means by which an intervention is transmitted so that it can affect outcomes. The *impact* domain describes how (e.g. in what direction and with what strength of association) the program's

intervention is expected to effect change in an outcome. Finally, the domains of *implementation environment* and *generalization* describe the context in which an intervention occurs and whether and how the impact of treatment on an outcome via the intervening mechanisms can be generalized to other contexts. Chen notes that "a systematic combination of all six domain theories constitutes a superordinate theory of a program…" (1990, pp. 51).

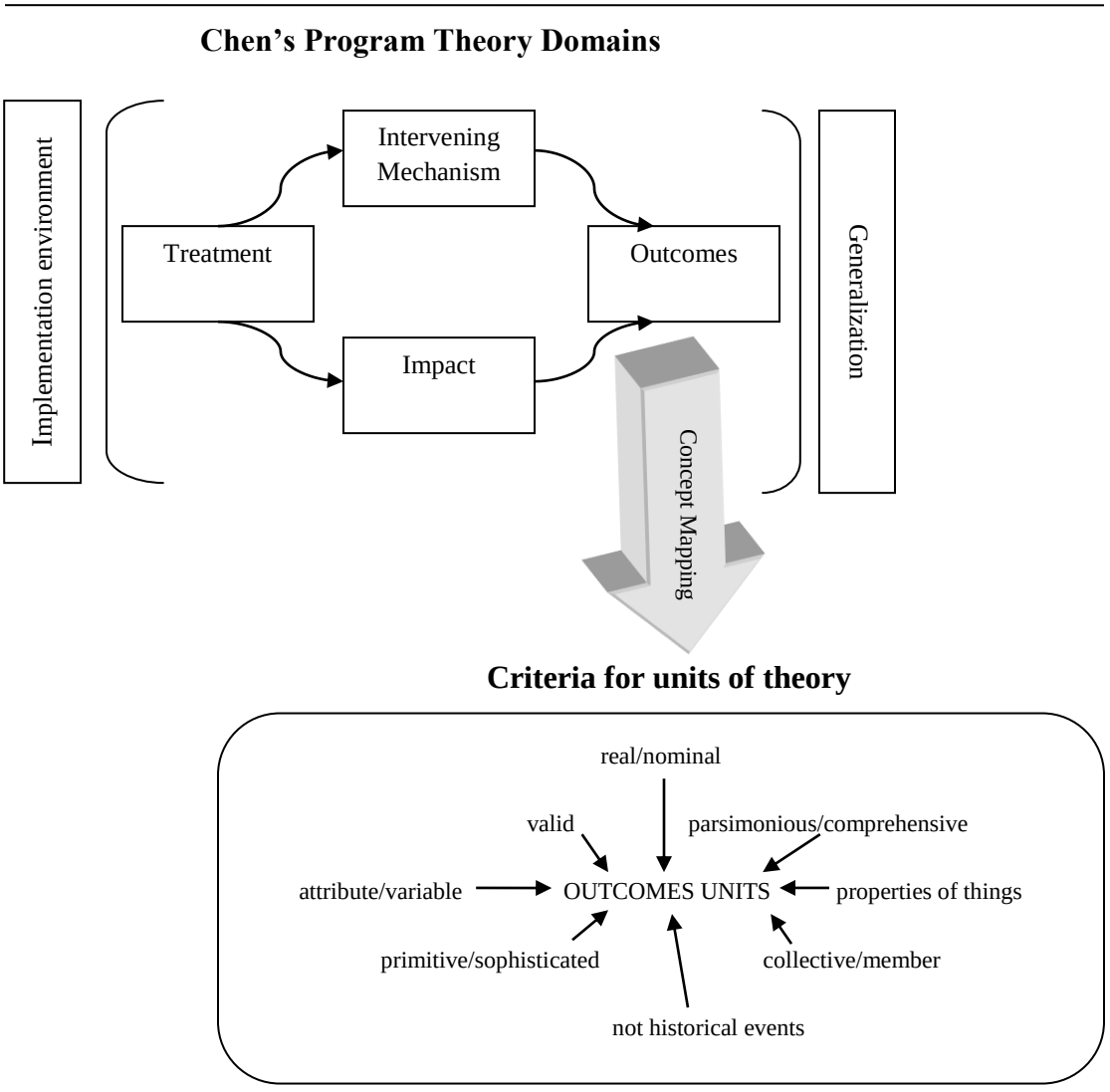


Figure 1. Relationship between Chen's (1990) program theory domains and the use of concept mapping for developing theoretical units (Dubin, 1978) in the outcomes domain.

In this study, concept mapping will be used to develop the outcomes domain for a community organizing intervention. As the figure and the research questions indicate, the study will use Dubin's (1978) criteria to explore how well concept mapping can be used to develop units which make up the outcomes domain. It will also explore whether the units of the outcomes domain are both parsimonious and comprehensive (Whetten, 1989) for that domain, and what evidence exists for the validity of the articulated outcomes.

Definitions

Definitions for the terms *theory*, *program*, *program theory* and *theory-based evaluation* are now proposed, as these ideas will appear throughout the study. First, consider a definition of *theory*. Chen writes: "Theory is a frame of reference that helps humans to understand their world and to function in it….theory provides not only guidelines for analyzing a phenomenon but also a scheme for understanding the significance of research findings" (1990, p. 17). Lynham (2002a, p. 221) states: "What is the purpose of good theory other than to describe and explain how things actually work and, in so doing, to help us improve our actions in this world?" And Patterson notes that: "…a theory is an attempt to organize and integrate knowledge and to answer the question 'Why?' " (1986, p. xix). Taken together, these ideas suggest the following definition for a theory. Theory is: an organized statement of assumptions and knowledge about a specified phenomenon which both (a) describes how it works and (b) explains why it works.

Next, consider definitions for the terms *program* and *program theory*. This study will follow Chen's understanding that a program "…is the purposive and organized effort

to intervene in an ongoing social process for the purpose of solving a problem or providing a service" (1990, p. 39). In other words, programs are one attempt at improving human beings' circumstances in the world. Extending the definition from above, one can understand program theory to be defined as follows: an organized statement of assumptions and knowledge about a specified program which both (a) describes how it should and does work, and (b) explains why it should and does work. This definition is consistent with Chen's notion that such a theory is two-fold. It has one portion which specifies "what the structure of a program *should* be" and another portion which states "…what *are* the underlying causal mechanisms that link the relationships among program treatments, implementation processes and outcomes…" (Chen, 1990, p. 43). Thus, a complete program theory lays out how the program both *should* and in fact *does* work.

Similarly, Weiss (1997b) writes that theories concerning programs are two-fold in nature, consisting of an implementation theory (how a program *should* work) and a programmatic theory (how it actually *does* work). The distinction between prescriptive/*should* and descriptive/*does* theory can also be connected to Argyris and Schoen's notions of "espoused" theory versus "theory-in-use" (1974; 1996, p. 13). An espoused theory is "…advanced to explain or justify a given pattern of activity" (p. 13). In contrast, theory-in-use is a theory that actually underlies an action. Theory-in-use "…must be constructed from observation of the pattern of action in question" (Argyris & Schoen, 1996, p. 13). All of these authors suggest that theory can describe and explain an ideal situation (i.e. how the activity or action *should* work) and can describe and explain an actual situation (i.e. how an activity or action in fact *does* work). Indeed, a complete

program theory must do both. The portions of program theory articulated in this study will be descriptive outcomes. Such descriptive theory is necessary before attempting to articulate intervening mechanisms which explain *why* outcomes occur.

Finally, to define *theory-based evaluation* one can turn to Weiss's early work (1972) which appears to be the first to suggest that a program's theory should be a key component of that program's evaluation (Weiss, 1997b; Worthen, 1996). In her book, Weiss discusses the fact that a model of the program's processes (i.e. the program's theory; see editor's note in Weiss, 1996/1972) can begin to shed some light on *why* certain outcomes occur or do not occur (Weiss, 1972, p. 51). This suggests that a theory-based evaluation stands in contrast to a more black-box type of evaluation (Chen & Rossi, 1987). In a black-box evaluation, it may be established that certain inputs co-vary with outputs or that there exists a statistical association of inputs with outputs. But a black box evaluation provides no further information about the nature of the association. A theory-based evaluation seeks to go further and explore *why* such an association occurs.

Philosophical Paradigm

Before continuing with a literature review and a description of the proposed study, the paradigmatic perspective for this research – critical realism – is described. Critical realism traces its intellectual roots to the work of Roy Bhaskar (Bhaskar, 1975, 1998) and it can be viewed as a "…middle ground between positivism and relativism" (Bechara & Van de Ven, 2007, p. 61). Bechara and Van de Ven provide an accessible introduction to critical realism. Ontologically, critical realism acknowledges that "…there is a real world out there (consisting of material, mental, and emergent

products)" and that such reality is layered, stratified, multi-dimensional and mind-independent (Bechara & Van de Ven, 2007, pp. 37, 64). Epistemologically, critical realist thinking accepts that humans' understanding of reality is "limited" (Bechara & Van de Ven, 2007, p. 37). Furthermore, Bechara and Van de Ven state that inquiry cannot be "impartial" and that, "…all facts, observations and data are theory-laden, implicitly or explicitly" (2007, p. 38). In other words, no theories or accumulated knowledge are objective in the sense of being independent of the observer who articulates them; all knowledge is interrelated with the perspective of the knower. Now consider the axiology of critical realism. An axiology of research refers to values that influence a researcher's choices in conducting inquiry (Guba & Lincoln, 2005). Van de Ven states the following assertion for critical realism:

> Most phenomena in the social world are too rich to be understood adequately by any single person or perspective….[and] any given theoretical model is a partial representation of a complex phenomenon that reflects the perspective of the model builder....this requires scholars to be far more reflexive and transparent about their roles, interests, and perspectives...than they have [been] in the past. (2007, p. 14)

Thus, an understanding of social phenomena as complex leads to a set of values for research. These values include using multiple perspectives and encouraging reflexivity regarding the researcher's point of view. Methodologically, critical realism is also pluralistic and inclusive. Similar to the paradigm's epistemological assumption, Bechara and Van de Ven note that "…no *form of inquiry* [emphasis added] can be value-free and impartial; each is value-full" (2007, p. 38). Thus, the paradigm recognizes that methods of inquiry reflect an underlying set of values and perspectives. The selection of method should depend on the research context because "…some methods are better warranted than others depending on the phenomenon" (Bechara & Van de Ven, 2007, p. 38).

Finally, consider the teleology of critical realism. Teleology answers the question "To what end?" is the research conducted (Lincoln & Lynham, 2011). In other words, what is the purpose of research? Bechcara and Van de Ven opine that "…science is an error-correction process that is based on evidence from the world, rather than merely reflecting the scientist's opinions of the world" (2007, p. 65). Further, the authors quote McKelvey (who cites Holton): "…the singular advantage of the realist method is its empirically-based, self-correcting approach to the discovery of truth" (Holton, 1993; McKelvey, 2002, p. 754). Thus, critical realist science pursues truthful knowledge that derives "…at least in part [from] …the way the world is" (Bechara & Van de Ven, 2007, p. 58).

Pawson and Tilley (1997) describe a *realist* perspective for evaluation which is also consistent with the *critical realist* view described above. First, they note that:

> Realism has sought to position itself as a model of scientific explanation which avoids the traditional epistemological poles of positivism and relativism. Realism's key feature is its stress on the mechanics of explanation, and its attempt to show that the usage of such explanatory strategies can lead to a progressive body of scientific knowledge. (1997, pp. 55-6)

As with Bechara and Van de Ven (2007), Pawson and Tilley stress that realism is an ontologically moderate paradigmatic framework, avoiding both completely positivist and completely relativist points of view. Pawson and Tilley also concur with Bechara and Van de Ven that realism incorporates a teleological end of knowledge accumulation. It is the stress on explanation and knowledge accumulation which is central to Pawson and Tilley's thinking about program evaluation from a realist perspective.

Pawson and Tilley provide a formula to express their idea about realist explanation: "…the basic realist explanatory formula [is]: *regularity = mechanism + context*" (1997, p. 56). In a related paper, Tilley explains this formula. He writes that

"…the realist understands causality in terms of underlying causal mechanisms generating regularities. The underlying causal mechanism will often be hidden....in the natural world, potential causal mechanism [*sic*] will only be activated if the conditions are right for them" (Tilley, 2000, pp. 4-5). Thus, causal explanations are not of a constant conjunctive nature where it is assumed that "like will always produce like" (Tilley, 2000, p. 4). Rather, realism allows for consideration of the effect of context when studying cause. Pawson and Tilley modify the explanatory formula for the purpose of enunciating program theory (which they deem critical to a good evaluation). Their modified formula for program theory is: *outcome = mechanism + context* (1997, p. 57). This three-part formula reflects all six of Chen's (1990) theoretical domains (see Figure 1 on page 5). Chen's outcome domain is what Pawson and Tilley seek to explain (*outcome*). Chen's treatment, impact and intervening mechanism domains all relate to explaining how and why a program will cause change in some outcome measure – these domains describe a program's *mechanism*. Finally, Chen's implementation environment and generalization domains relate to the *context* of a specific program and the circumstances in which any conclusions from an evaluation might be applicable. Thus, Chen's theoretical framework accommodates Pawson and Tilley's realist explanatory formula quite well.

As shown, realism provides a framework to study cause-and-effect-in-context. The ability to deal with causal outcomes is important for program evaluation (Cook, 2000; Rogers & Weiss, 2007). Thus it is vital to begin building a theory that accommodates cause-and-effect thinking. Although the current study will not deal explicitly with causal mechanisms (this would be part of the intervening mechanism domain as named by Chen, 1990), it will lay the groundwork for a complete program

theory of community organizing which can accommodate causal explanations. As a precursor to causal thinking, however, clear outcomes must be articulated. In Chapter 3, Metro Organizations for People (MOP) – the organization which provides context for this study, is introduced. MOP's executive director expressed one of his wishes for a comprehensive program theory as follows: he'd like to know how MOP members are different from "Joe on the street" (M. Kromrey, personal communication, November 10, 2009). In other words, Kromrey would like to understand whether MOP's outcomes include personal changes for MOP's own members. This comment implies an assumption on his part that MOP's work does cause its members to be different in some sense from people who do not participate in the community organization. Undoubtedly, MOP's funders would also like to understand what outcomes are caused by MOP's work. Using a realist perspective for theory-building will allow both the acknowledgment of multiple participant and stakeholder perspectives and the future inclusion of cause-and-effect thinking in the program theory.

Delimitations and Limitations

The current study is delimited in two appreciable ways. First of all, due to time and resource limitations, it will provide evidence for the utility of concept mapping in theoretical unit articulation for only one of Chen's six domains of program theory. Further research would be required to explore whether or not concept mapping is useful for articulating theory in some or all of the remaining five domains. This delimitation also implies that the study will explore the utility of concept mapping for theory-building only in regards to the descriptive portion of a program theory and not the causal-explanatory portion of such a theory. Inherent in this delimitation, however, is a future

opportunity to extend the work begun here to other theory domains, including the intervening mechanism domain which articulates causal links. Secondly, the study is limited in the sense that it will work with only one program and one group of participants at one community organization in one city. After a single study, it will not be possible to generalize results in the sense suggested by Chen of applying them "…to future pertinent circumstances or problems in which stakeholders are interested" (Chen, 1990, p. 65). In this case, "future pertinent circumstances" would be whether or not concept mapping can support theory articulation for other complex social interventions. Further research in the context of other social interventions would be needed to determine whether it is appropriate to generalize or "transfer" (Lincoln & Guba, 1985, p. 217) the methodological results of the study to other situations.

Significance

This study is significant in three ways. First, the theory-based evaluation literature lacks methodological examples of how one might evaluate the quality of articulated theory. The proposed study will use several of Dubin's (1978) and Whetten's (1989) criteria to evaluate how well concept mapping works to develop an outcomes domain for program theory. Second, while answering questions about validity, the study explores expansion of the type of statistical clustering methods used for Trochim's concept mapping research tool (1989). Finally, the study lays groundwork for longer-term research aimed at building a program theory for grassroots community organizing. At this time, a comprehensive program theory for community organizing (encompassing all six of Chen's domains) is absent from the literature.

Chapter 2 – Literature Review

As discussed in chapter 1, this study seeks to contribute to discussions in the literature on how to improve theory-based evaluation, and, in particular, it seeks to expand methodological discussions around how researchers might better articulate program theory. To build the case for such a study, this chapter first provides an overview of discussions from a more general literature on theory-building, which can in turn inform discussions about methods for program theory development. Secondly, the chapter reviews literature on the history and current "state of the art" in theory-based evaluation. Finally, it provides an overview (from a variety of fields in the social sciences) of the use of the concept mapping research tool, including several attempts to connect concept mapping with theory-based evaluation.

Theory-Building

Weiss (1997a) suggests that one fundamental challenge to TBE is the difficulty in constructing a program theory. If one is to develop or articulate a theory, one needs a general process or methodology by which to do so and one also needs specific tools to use. To understand methodology for theory-building, one can reach beyond the evaluation literature into the literatures of other applied fields such as human resource development and psychology. It is in these fields that one finds the expertise necessary to understand how to go about building a theory.

Lynham's general method for theory-building. Lynham's (2002a) general theory-building method for applied disciplines provides a systematic framework for the

entire process of building a theory. Though Lynham's context and references are primarily from the field of human resource development, she has positioned her methodology for theory-building as a general one for applied fields. An extension to the area of theory-based evaluation is thus appropriate. Lynham's general method is "…a recursive system of five distinct phases" (2002a, p. 229) including: (1) conceptual development, (2) operationalization, (3) confirmation or disconfirmation, (4) application and (5) continuous refinement or development. Conceptual development includes formulation of "…initial ideas in a way that depicts current, best, most informed understanding and explanation of the phenomenon, issue, or problem in the relevant…context" (Lynham, 2002a, p. 231). Then, operationalization "…reaches toward an overlap between the theorizing and practice components of the theory-building" (2002a, p. 233). This step frequently consists of expressing results from the conceptual development phase in terms of "…confirmable propositions, hypotheses, empirical indicators, and/or so-called knowledge claims" (Cohen, 1991; Lynham, 2002a, p. 233). Confirmation or rejection occurs when the theory-builder executes an appropriately designed research study intended to confirm or disconfirm the operationalized expression of the conceptually developed theory. Application follows confirmation or disconfirmation and it involves the use of a developed, operationalized and confirmed theory to inform a specific problem. This enables use of "…experience and learning from the real-world application of the theory to further inform, develop, and refine the theory" (Lynham, 2002a, p. 233). Finally, Lynham notes that a theory is never complete (see also Cohen, 1991; Root, 1993). Thus, the fifth and last step ties the rest together and requires that the theorist continually repeat the other steps in a cyclical

process in order to "…ensure that the theory is kept current and relevant and that it continues to work and have utility in the practical world" (Lynham, 2002a, p. 234).

Dubin's method for theory-building. Dubin (1978) outlines a method for theory-building which can be considered a "specific variation of …[Lynham's] general method of theory-building research in applied disciplines" (Lynham, 2002a, p. 242). There are eight steps in Dubin's method: (a) specify units, (b) specify laws of interaction, (c) specify boundaries, (d) specify applicable system states, (e) develop propositions, (f) find empirical indicators of key terms, (g) form hypotheses and (h) test theory (Lynham, 2002b). Relying on Lynhams' (2002b) description of Dubin's method, here follows an extremely brief overview for each of Dubin's eight steps to build a theory. First, the units of the theory can be understood as the basic concepts or ideas which make up the theory. Specifying these is a first step. Then, laws of interaction must be specified which "…describe the interaction among the units of the theory" (Lynham, 2002b, p. 249). In the next two steps, the theory-builder articulates (a) boundaries, indicating where a theory can be expected to apply, and (b) system states, specifying the "…condition[s] under which the theory is operative" (Lynham, 2002b, p. 256). In step five, one finds that propositions are developed which are "truth statements about the theory" (Dubin, 1978, p. 160; Lynham, 2002b, p. 261). These are *not* empirically validated statements, but rather, statements which "…are logically derived from the theory itself" (Lynham, 2002b, p. 261). Next, empirical indicators are found for each unit. These indicators are the means by which measurements are recorded for a unit included in the theory. Seventh, the theory-builder must construct hypotheses in order to test the theory. Dubin notes that multiple hypotheses may issue from any one of the

logical propositions of the theory (Dubin, 1978, p. 208). Finally, the theory specified using steps 1-7 is tested through a "developed plan of research" (Lynham, 2002b, p. 269). Further detail on Dubin's method is available in Lynham (2002b) and in Dubin (1978).

The articulation of an outcomes domain (Chen, 1990) for a program theory takes place during step one, unit specification, in Dubin's (1978) method. More generally, domain articulation can also be considered a part of the conceptualization phase of Lynham's general method (2002a). The study hypothesizes that Chen's outcomes domain must contain several units (i.e. theory building blocks, Dubin, 1978). If these units are well-specified, they can be used in constructing a larger program theory. In Chapter 3 of his text, Dubin discusses several "important distinctions between paired characteristics of units" (p. 37) which should be considered while units are being specified in step one. These paired characteristics are: concept and unit, thing and "property of thing" (p. 40), unit and event, attribute and variable, real and nominal, primitive and sophisticated and collective and member. Dubin's implication is that units for theory must conform to these characteristics. Next, Dubin's distinctions are reviewed in more detail, as well as Whetten's (1989) discussion of parsimony and comprehensiveness. Taken together, Dubin and Whetten provide a basis for establishing criteria that can evaluate whether concept mapping is a good tool for program theory specification.

Two of Dubin's (1978) several paired distinctions articulate required criteria for *bona fide* units of a theory. Consider first the "thing versus property of thing" distinction. Dubin writes, "…when we consider classifying units of a theory in the behavioral sciences…we build our theories about the properties of things rather than the things

themselves. We focus our theories upon selected characteristics of objects rather than upon the objects" (1978, p. 40). To illustrate this point, Dubin gives the example of studying individuals versus the morale of individuals. Behavioral scientists do not study or theorize about individuals per se, but they may do so regarding individuals' morale. Thus, theoretical units should describe properties of a thing (e.g. properties of individuals, social groups or occurrences, such as morale) rather than the thing itself.

In addition, the theorist must also ensure that a unit does not describe an event. "The distinction [between unit and event] rests on the question of number. An event happens only once…" (Dubin, 1978, p. 42), whereas a unit must be plural. The example given to illustrate this point is about war. Dubin notes that the American Civil War is an event, whereas "all the wars of the United States" (p. 43) can function as a unit for theory-building. This helps distinguish theory from historical explanation. Thus, for a potential unit to be useful for theory-building in Dubin's method, it must not be a one-time historical event.

Dubin (1978) then presents four more distinctions relevant to unit specification. For these distinctions, a unit of theory is required to display either one or the other of the characteristics in the pair. The pairs are: attribute or variable, real or nominal, primitive or sophisticated and collective or member.

First of all, a unit *can* be either an attribute or a variable, but it *must* be one or the other. An attribute "…is a property of a thing distinguished by the quality of being present" whereas a variable "…is a property of a thing that may be present in degree" (1978, p. 44). Membership in a specific political party is an attribute unit (either one is a member or one is not) whereas intelligence is a variable unit. In the case of intelligence,

one can have a greater or lesser degree of this property; intelligence is not exclusively either present or absent.

Secondly, the real versus nominal distinction refers to whether an empirical indicator can be invented or found for the unit. (Finding empirical indicators is the sixth step of Dubin's complete eight-step theory-building process.) A unit is real if there is a reasonable probability of finding or inventing an empirical indicator to stand for the unit. It is nominal if no such empirical indicator can at present be found. Note that only the portions of a theory which contain real units can be empirically tested. However, Dubin suggests that including nominal units in a theory is valuable because it helps to extend the boundaries of scientific knowledge. Examples of such nominal units from the social sciences are the "id" and "ego", "power", "charisma" and "society" (Dubin, 1978, p. 52).

Next, Dubin considers primitive versus sophisticated units. Simply put, a sophisticated unit is defined and a primitive unit is not defined. An example of a primitive unit occurs when the scientist observes that there is some unknown X needed to account for an observation. As an example, Dubin writes:

> When Dr. [Alexander] Fleming was confronted with some odd stuff under his microscope…he then said, in effect, there is some *X* here that is somehow connected with what appears in the field of the microscope. When his attention turned to explicating this [unknown] *X*, we were given the gift of penicillin. (1978, pp. 53-54)

A primitive unit in a theory is by nature temporary; the scientist attempts to define it so that the primitive unit may be changed to a sophisticated one.

Finally, there is the distinction between collective and member units. A collective unit is a property of a group of things. In contrast, a member unit is a property of just one thing. A social group's degree of stability could be considered a collective unit, while an

individual's level of education could be considered a member unit. This last pair concludes the discussion of Dubin's (1978) four either/or distinctions. Note that as units are discovered or defined the theorist *must* be able to make four two-way classifications describing the nature of the unit: attribute or variable, real or nominal, primitive or sophisticated and collective or member. If such classifications cannot be made, the units are not useful for building theory using Dubin's framework.

Whetten: building blocks for theory development. Whetten (1989) also sheds some light on criteria which may be used for evaluating the goodness of theoretical units. He writes that a complete theory requires four elements: What, How, Why and Who/Where/When. In Whetten's framework, the "what" (p. 490) of a theory are variables, concepts or constructs (collectively called *factors*), which are used to characterize a phenomenon of interest. The "how" (p. 491) of a theory is the set of relationships that connect variables, concepts and constructs with each other. "Why" (p. 491) provides an explanation for the selection of factors and the relationships connecting them. Finally, who/where/when (p. 492) provide the context description and boundaries for the theory. Whetten's "what" factors can be considered analogous to Dubin's (1978) units. Whetten states that "…two criteria exist for judging the extent to which we have included the 'right' factors: *comprehensiveness* (i.e. are all the relevant factors included) and *parsimony* (i.e. should some factors be deleted because they add little additional value to our understanding)" (1989, p. 490). Thus, in addition to Dubin's distinctions, Whetten's criteria of comprehensiveness and parsimony can also reasonably be used to judge the adequacy of theoretical units.

Trochim and Leeuw: program theory. If one applies Lynham's (2002a) framework to the problem of articulating a program's outcomes domain, one begins to build a program theory using the conceptual development phase of the general model. Dubin's unit articulation is an example of such conceptualization (Lynham, 2002b). Lynham (2002a) writes that conceptualization includes articulation of the theory builder's best and most informed understanding of various program-related phenomena. Trochim (1998) argues that the best source of program theory is the people who know the program well and that evaluators articulating theory should: "…include the implicit theories of the people closest to the program" (1998, p. 246). One might therefore ask how researchers or evaluators can better understand the thinking of those who know the program well in order to execute the conceptual development and unit articulation which are needed for theory-building.

Leeuw (2003) provides a possible answer. He tackles the problem of what methods a researcher might use to conceptualize a program theory, though he uses the term "reconstruction" (p. 6) of the program's theory instead of "conceptualization". He writes that there are three methods for reconstructing program theories: a policy-scientific approach, a strategic assessment approach and an elicitation approach. The policy-scientific approach is a six-step process which Leeuw characterizes as connected with mainstream evaluation methodologies and which relies on "interviews, documents, and argumentational analysis" (2003, p. 18). The strategic assessment approach is somewhat dialectical and "has strategic assessment, group dynamics, and dialogue as its core" (2003, p. 18). Finally, the elicitation approach relates to ideas from cognitive and organizational psychology. Leeuw notes that Trochim's concept mapping method (1989)

is an example of the elicitation approach. This study explores Leeuw's claim by asking whether the output of a concept mapping exercise can be used to build theory. Before turning to a discussion of concept mapping, however, the next sections review literature recapping the history and development of theory-based evaluation.

Theory-based Evaluation

Development of theory-based evaluation. If one wanted to attribute the original idea for what is now called theory-based evaluation (TBE) to a single person, that person would be Carol H. Weiss, an emerita professor in the Graduate School of Education at Harvard University. Weiss herself notes that Edward Suchman made one of the earliest known references to the idea of a program's theory (Suchman, 1967; Weiss, 1997b). But it is Weiss's book entitled *Evaluation research: Methods for assessing program effectiveness* (1972) which appears to be the first text suggesting that a program's theory should be a key component of that program's evaluation (Weiss, 1997b; Worthen, 1996). In the book, Weiss notes that the program's theory "link[s] the events of the program to the desired effects" (1972, p. 49). She also discusses the fact that a model of the program's processes – what is now called a program theory – (see editor's note in Weiss, 1996/1972) can begin to shed some light on *why* certain outcomes occur or do not occur (Weiss, 1972, p. 51). This type of thinking stands in contrast to a more black-box type of evaluation (Chen & Rossi, 1987) in which it may be established that certain inputs covary with outputs, but which includes no discussion of *why* this might be the case.

Although one might credit Weiss with some of the earliest thinking about theory-based evaluation, Worthen (1996) also notes several other not-oft-cited articles from the 1970s in which ideas about theory-based evaluation were developed. These articles

include Fitz-Gibbon and Morris's (1996/1975) early article, which focuses on *how* a program might produce specific outcomes and which suggests that such conclusions should drive the design of an evaluation. In addition to Fitz-Gibbon and Morris's thoughts from the 1970s, Worthen writes that Bickman (1979) supported theory development by social psychologists and that Quay (1979) proposed the idea that "…valid causal attributions about programs depend on knowledge of program philosophy and implementation…" (Worthen, 1996, p. 170). Finally, Wholey's (1979) work on "evaluability assessment" (p. 17) completed a trio of ideas which characterize some of the influences upon which later thinking about theory-based evaluation has been built (Worthen, 1996).

Worthen also writes that, by the mid-1980s, theory-based evaluation had evolved into a "powerful movement in program evaluation" (1996, p. 169). Both Worthen (1996) and Weiss (1997b) note that much TBE work in the early 1980s was done by Huey-tsyh Chen and Peter Rossi (1980, 1983, 1987). In an early article, Chen and Rossi (1980) note that many social programs are found to be without effect. They go on to note that this result may be due to true ineffectiveness (though it is hard to believe that almost all programs in almost all places achieve no outcomes), or it may be due to the fact that the methods used to evaluate social programs (i.e. scientific experimental and quasi-experimental methods) are inappropriate. Chen and Rossi (1980, p. 106) suggest that a "theory-driven" approach can assist with evaluation by focusing on a wider range of outcomes for the program than those outcomes which were targeted by the policy-makers who developed the program. As with earlier authors, Chen and Rossi (1980) focus on the importance of a model for the program and on the importance of examining intervening

processes which link the program treatments and outcomes. They conclude by suggesting that there are several advantages to a theory-driven evaluative process, including (a) greater opportunities for detecting non-zero program effects, (b) more information for policy- and decision-makers and (c) the opportunity for evaluation work to contribute to more general social science theories. In a later article, Chen and Rossi (1987) argue that TBE offers the benefit of being able to synthesize competing approaches to validity in research (i.e. internal versus external). They write that "the central argument in our approach is that a model or theory should be formulated in a program evaluation and the modeling process should include the identification of potential threats to validity in research" (Chen & Rossi, 1987, p. 102). In other words, TBE may have some potential to resolve validity trade-offs and offer a way to design studies where one type of validity is not completely sacrificed to achieve another.

Despite more developed thinking about theory-based evaluation during the 1980s, it was apparent that, despite its promise, TBE was not being used very often. Lipsey, Crosse, Dunkle, Pollard and Stobart (1985, p. 7) conducted a general review of the state of the "art" and "science" of evaluation. As part of their review, they classify the sophistication of program theory (presented in the 122 studies they reviewed) according to five levels: black-box, program strategy, program principles, hypothesis testing and integrated theory. Only 9% of the studies they examine had articulated an integrated theory, which the authors define as:

> …an a priori theory within which the specific formulation of program elements, rationale and causal linkages was embedded. The program theory was more general than the specific application represented by the program under investigation and was derived from some source other than the experience or folk wisdom of the program personnel or the evaluation researchers. (1985, p. 22)

The remaining studies in the Lipsey, et al, review either lacked any type of program theory or did not articulate a comprehensive theory. Bickman also notes many barriers to the use of program theory and that a "…careful examination of the field of evaluation indicates that few evaluations are theory based" (Bickman, 1989, p. 387). He suggests that changes in graduate training and journal policies, as well as better methods for theory development might increase the use of TBE (Bickman, 1989).

Reflection on theory-based evaluation continued into the 1990s. Chen expanded on many of the ideas he and Rossi published during the 1980s when he produced a book-length treatment of TBE entitled *Theory-driven evaluation* (1990). This text provides a more in-depth description of what a program theory should be; this text is used in the current study as a foundation for understanding the elements of program theory. Chen includes the idea that such a theory should be two-fold with a "prescriptive" portion, specifying "what the structure of a program *should* be" and a "descriptive" portion, stating "…what *are* the underlying causal mechanisms that link the relationships among program treatments, implementation processes and outcomes…" (Chen, 1990, p. 43). Further, a good program theory would consist of the six previously mentioned domains: treatment, implementation environment, outcome, impact, intervening mechanism(s) and generalization. Chen also offers thoughts about perspectives or paradigms which a researcher or evaluator can use when constructing a program theory and he discusses the different types of evaluations which may flow from the six domains.

Also during the 1990s, Carol Weiss published the second edition of her 1972 text. It was re-titled *Evaluation: Methods for studying programs and policies* (1998). In her preface she notes that she has expanded the "…few pages on program theory in the

earlier book into a full chapter" (Weiss, 1998, p. x), indicating her continuing belief of the importance of program theory to good evaluation. Toward the end of the decade, Weiss provides a more detailed reflection on the continuing challenges to using theory-based evaluation (1997a). Weiss characterizes TBE as a "plausible and cogent concept" which has nevertheless been very "slow in coming into practice" (1997a, p. 501). Weiss notes a dozen such challenges, which are: (a) unclear program theory, (b) confusion over the difference between process and outcome in program theory, (c) multiple possible theories for one program, (d) difficulty in constructing program theory, (e) lack of clarity on the difference between black-box process-outcome evaluation and theory-based evaluation with explicitly identified links between process and outcome, (f) large time and resource requirements for TBE, (g) continued challenges with measurement error, which also affect social science in general, (h) demanding analysis of both qualitative and quantitative data, (i) a lack of generalizable results, (j) a need for counterfactual evidence to make causal inference, (k) a tendency to avoid evaluating fluid or changing program goals and (l) difficulties with testing theories. It is a disappointingly long list.

Most recently, the discussions about theory-based evaluation have become more methodological. Several articles appear in a special issue of the journal *New Directions for Evaluation*, edited by Patricia J. Rogers, Timothy A. Hasci, Anthony Petrosino and Tracy A Huebner. In this volume, Cook (2000) asserts strongly that theory-based evaluations cannot by themselves provide causal inference; they can only do so if they incorporate a counter-factual for comparison. Davidson (2000), on the other hand, argues for alternative understandings of causation. Petrosino (2000) discusses possible connections between theory-based evaluation and meta-analysis. In the *American*

Journal of Evaluation, Leeuw (2003) provides a discussion of methods for articulating theory and notes that program theory must deal with *underlying mechanisms* and not just be presented as a flowchart.

Finally, in a recent article, Rogers (2007) repeats several points that Weiss made in her 1997 articles (Weiss, 1997a, 1997b). Rogers notes that program theories used for evaluation are often of poor quality and also that "…the ways program theory [*sic*] are used to guide evaluation are often simplistic" (Rogers & Weiss, 2007, p. 65). She echoes Weiss's concern about a lack of clarity between process and outcome. And Rogers comes full circle from the beginnings of TBE when she reiterates that a program theory should be more than a theory of implementation which enumerates activities and outcomes. It must also deal with causation and the *underlying mechanisms* which produce those outcomes. A diagram with boxes and arrows may not be enough (Rogers & Weiss, 2007). Rogers observations bring us back to Weiss's original idea that a program's theory must "link the events of the program to the desired effects" (1972, p. 49). After forty years of trying, this is still a very difficult thing to do.

Strengths of theory-based evaluation. Although strengths of theory-based evaluation have been hinted at in the historical discussion above, they are here reviewed before turning to look at some empirical examples of TBE which display these strengths (and some weaknesses, too). First, theory-based evaluation provides a more holistic approach to evaluation, looking at the why and how (Fitz-Gibbon & Morris, 1996/1975; Weiss, 1972) of a program. TBE provides a deeper picture of the workings of a program than the picture which results from examining only how a set of program activity

measurements co-vary with a set of desired outcomes. Such a picture is useful for purposes of improving a program.

A second strength is that theory-based evaluation provides opportunities for theory-building, not just evaluation of an individual program (Chen & Rossi, 1980). Weiss suggests a similar idea when she notes: "My long-range hope is that evaluation will not only be based on theory but [will] also *contribute* to the cumulation of theoretical knowledge" (1997b, p. 52). This suggests a role for TBE not just in evaluating the application of knowledge but also in the building of basic knowledge itself (Chen & Rossi, 1980). As this literature review suggests, connecting program theory development with more general methods for theory-building such as Lynham (2002a) and Dubin (1978) may further the connection between theory-based evaluation and the cumulation of knowledge.

Finally, theory-based evaluation lets research questions drive the method instead of the other way around (Chen & Rossi, 1987). This means that TBE makes it more likely that the method chosen for evaluation will be appropriate to the program and its theory, instead of the research method being chosen a priori. Letting method follow research questions has frequently been recognized as good research practice (see Johnson & Onwuegbuzie, 2004), but it can be easy to let the methods drive the questions instead.

Challenges of theory-based evaluation. Weiss's recent enumeration of the many challenges faced by theory-based evaluation (Weiss, 1997a, see above) provides a comprehensive list that does not require any expansion. They are all summarized in the above historical discussion. Of the challenges she lists, however, the most fundamental two are: (a) difficulty in constructing program theory and (b) lack of clarity on the

difference between black-box process-outcome evaluation and theory-based evaluation with explicitly identified links between process and outcome. These two are also the challenges which are most specific to TBE as opposed to affecting applied social science in general. The second challenge is fundamental because if researchers and evaluators do not understand what theory-based evaluation really is, they should not expect to do it well. The first challenge is also fundamental; if researchers or evaluators cannot construct a robust program theory then they have no framework for theory-based evaluation and no mechanism by which to interpret evaluation findings. It is the first fundamental challenge – that of constructing or articulating program theory – which this study seeks to address.

Empirical examples of theory-based evaluation. Birckmayer and Weiss (2000) review six recent studies that use theory-based evaluation. These studies are all in the field of health promotion, which the authors note has seen wider use of TBE than have other fields of practice. The six studies include a nutrition education program (Brug, Steenhuis, Van Assema, & De Vries, 1996), two anti-smoking programs (Flay, et al., 1995; Murray, Prokhorov, & Harty, 1994), a sex education program (Eisen, Zellman, & McAlister, 1992), a heart disease prevention program (Puska, et al., 1985) and an alcohol-related injury prevention program (Holder, Saltz, Treno, Grube, & Voas, 1997). All of these studies have relatively strong experimental or quasi-experimental designs. In addition, Birckmayer and Weiss note that the authors for each article articulate a theory for how they expect the intervention to lead to a desired outcome (or outcomes). After briefly describing the studies, Birckmayer and Weiss note that "although each of the evaluations has some modicum of theory involved, the authors are not always explicit

about what they learned from TBE over and above what they would have learned without it" (2000, p. 423). Thus it appears that the given theories provide a slightly more holistic picture of the programs, and they offer some information about the "how" and "why" of an intervention's impact on desired outcomes. However, it also appears that a complete theory is not really central to the six evaluations and that the studies do not fully capitalize on the opportunity to move away from a more limited black-box process-outcome evaluation.

In contrast to the studies described by Birckmayer and Weiss (2000), Carvalho and White (2004) provide a strong example of theory-based evaluation in the area of international development. They articulate a theory about social funds and how the money disbursed from these funds can benefit poor communities. These authors preceded their evaluation with development of a program theory for social funds. They note that "this background work [i.e. program theory development] helped to guide the analysis of World Bank social fund project documents…and *resulted in identifying the assumptions to be tested, the data to be collected, and the instruments to be used for data collection* [emphasis added]" (2004, p. 146). First, these authors articulate a general theory of social funds using a page-long, textual description and a logic model (2004, pp. 143-145; Frechtling, 2007). Their theory encompasses the elements of Chen's (1990) framework. The treatment domain is represented by the establishment and operation of a social fund. The outcome domain is "sustainable benefits for the poor" (2004, p. 145). The implementation environment and generalization domains are not mentioned in the logic model, but they are nevertheless described beforehand in the text. The implementation environment consists of various sites around the world where social

funds are set up with at least partial financing and oversight from the World Bank. The generalization domain is closely linked to the implementation environment; the authors do not seek to generalize beyond the 66 projects in 40 countries which were supported with World Bank funding at the time of the evaluation. Intervening mechanisms, by which program designers expect the social fund (treatment) to effect sustainable benefits for the poor (outcomes), are also shown in the model. These include promotion of the fund to the local community, submission of community project proposals and several mechanisms designed to ensure continued community and governmental support for projects. Interestingly, an "anti-theory" (p. 145) is also articulated which posits several reasons why critics believe a social fund program might fail or have negative outcomes.

At this point in the paper, Carvalho and White (2004) have described five of Chen's six domains. They next describe anticipated impacts in two areas that will be specifically addressed in their evaluation: subproject sustainability and institutional development impact. To articulate these portions of the program theory, the authors include three detailed tables which summarize impacts to be measured or observed. Finally, empirical findings are presented in narrative form, and these findings make reference back to the articulated program theory. Carvalho and White's evaluation capitalizes on some of the strengths of theory-based evaluation which have been discussed in the literature. It provides a deep understanding of the workings of the program. And, because it both articulates a theory and looks for confirmation of that theory, the study can make a contribution to more general theory-building (Lynham, 2002a). This study does not fall prey to the challenges of poorly articulated theory and a lack of focus on causal links. More than one third of the paper is devoted to a discussion

of the program's theory and that theory includes examples of explicating the causal links between stages of the program.

Taut, Santelices, Araya and Manzi (2010) also provide a good example of program theory development in their study of the national evaluation system for teachers in Chile. The strengths of program theory explication in this study are two-fold. First, they involve several groups of stakeholders in the process, including the Ministry of Education, the Chilean teachers union, a local authorities association and the university-based program implementers. Multiple stakeholders can contribute to the development of a more comprehensive program theory (see Chen, 1990, for a related perspective on multiple stakeholders and program theory). Second, in the interviews which occurred as part of the program theory articulation process, the researchers explicitly tried to uncover the causal links between inputs and outcomes. They note, however, that they were only partly successful in this endeavor. Still, the fact that researchers deliberately tried to examine causal change (i.e. intervening) mechanisms in the development of a program theory is a first step toward addressing one of the major TBE challenges identified by Weiss (1997a).

Concept Mapping

Brief overview of concept mapping. Concept mapping was first presented as a cohesive research tool more than 20 years ago in an article by William Trochim (1989). In the article, Trochim suggests that concept mapping is an example of a structured conceptualization process. Such processes, through a series of procedurally-oriented steps, produce a collective representation of some concept or idea (Trochim & Linton, 1986). Concept mapping is a specific conceptualization process which may be

particularly helpful for program planning and evaluation. Trochim (1989) describes concept mapping in six steps: (a) preparation, which includes selection of participants and determination of the focus statements for brainstorming and rating, (b) generation of [response] statements via brainstorming, (c) structuring of statements via sorting and rating, (d) representation of statements by computing a concept map, (e) interpretation of maps and (f) utilization of maps. A group of participants is chosen for the process, and this group participates in a structured brainstorming process which includes generating, sorting, and rating a set of statements related to the topic in question. The sorted and rated statement set constitutes the data for a concept mapping study. A researcher or consultant then uses statistical tools to represent the dataset pictorially as a two-dimensional, clustered map of statements from the dataset. The final step in using concept mapping is interpretation of the meaning of the clustered map, which is often accomplished by conducting another session with the original participants. As Trochim (1989) suggests, the resulting map may be used for planning or program evaluation. Other authors have suggested that concept mapping may be used for program theory articulation (Leeuw, 2003; Rosas, 2005; Yampolskaya, Nesman, Hernandez, & Koch, 2004).

Examples of concept mapping in social research. Concept mapping has been used as a research method in a variety of fields of social inquiry. For example, in a marketing application, Bigne, Aldas-Manzano, Kuster and Vila (2002) used concept mapping to uncover the determinants of customer loyalty in the travel agency sector of the marketplace. Focus groups were convened in which participants "…were asked to indicate what motives would cause them to be loyal to a travel agency" (Bigne, et al.,

2002, p. 90). Sixteen response statements were generated and were subjected to the concept mapping process outlined by Trochim (1989). In this study, the number of clusters on the map were selected by targeting that number of clusters which would explain at least 60% of the variance in the data and where adding an additional cluster would increase by less than 5% the percentage of variance explained. This resulted in a three-cluster solution. The authors found that "in general, purchase customer loyalty is closely related to price, establishment [i.e. agency] attributes and relationship marketing" (Bigne, et al., 2002, p. 92).

Ridde (2008) used concept mapping for a portion of a health policy study conducted in Burkina Faso in West Africa. She notes that the purpose of the study was to understand people's values related to a health policy question, namely, "…making sense of the fact that…excluding indigents from access to [health] care was not perceived as a public issue" (Ridde, 2008, p. 8). Concept maps were constructed with two groups of participants, a group of nurses (in French) and a group of village health committee members (in the local language, Moore). The nurses responded to this brainstorming focus statement: "In Burkina Faso today, I think the notion of social justice means that...." (Ridde, 2008, p. 3). The Moore-speaking group of village leaders responded to a similar prompt in which the phrase "social justice" was translated as "…to not infringe upon others" (Ridde, 2008, p. 3). The response statements generated by the French-speaking nurses formed nine clusters: good governance, respect of human rights, justice and social peace, rational and efficacious management of aid, community participation, fighting against poverty, equitable and rational management of resources, social security and equity of access to basic social services (Ridde, 2008, p. 5). The statements by the

village health committee members also grouped into nine clusters: trust each other, to have integrity, honesty, mutual help, need others, transparency is good, the truth is coming, mutual support and come spontaneously to carry help (Ridde, 2008, p. 6). Because Ridde's article is methodologically focused, the author refers readers to another article for "…a comparative discussion of the results, as well as of the convergent and divergent elements of the two groups of actors" (Ridde, 2008, p. 8).

Finally, Bedi and Alexander (2009) used concept mapping in the context of a counseling study. Like Ridde (2008), they focused on methodological considerations for a large part of the article. One benefit mentioned several times is that concept mapping gives greater voice to the clients' perspectives and understandings of the counseling process than do many other traditional research methods. Bedi and Alexander interviewed forty counseling clients to generate response statements. A subgroup of thirty-one of the original participants sorted seventy-four statements which were used in the concept mapping process. The final map contained thirteen clusters, which were labeled as follows: "…Office Environment, First Impressions, Body Language, Listening Skills, Unconditional Positive Regard, Encouragement, Challenging, Ethics and Boundaries, Education, Referrals and Recommended Materials, Client Commitment, Procedural Clarifications and Choosing a Counselor" (2009, p. 86). Bedi and Alexander close their article with a comparison between concept mapping and another research method called the Critical Incident Technique (CIT). They conclude that CIT results are perhaps prone to undue levels of influence by researchers and that research might benefit from combining the best of both methods: rigorous data collection processes from CIT and statistical analysis from concept mapping.

Examples of concept mapping for theory articulation. As noted above, Leeuw (2003) considers Trochim's (1989) concept mapping process an example of an elicitation method for articulating program theory. A few studies have attempted to use concept mapping to elicit or articulate a program theory. However, the end result in these studies was not a program theory which was as well or as completely articulated as it could be.

Yampolskaya, Nesman, Hernandez and Koch (2004) used a concept mapping process (Trochim, 1989) to display staff conceptions of a Florida agency's program activities which are related to children's mental health services. The resulting concept map was then used to develop some portions of a logic model for the mental health program. The portions of the logic model which were derived from the concept mapping output were descriptions of (a) the program's target population and conditions, (b) the program's activities organized by category of services and (c) the program's strategies. The portions of the logic model dealing with outcomes were developed using interviews and discussion with program staff (Yampolskaya, et al., 2004, p. 194). Although the logic model in the paper provides a pictorial overview of the program and contains some information about treatments and outputs, it does not clearly explicate the links between treatments and outputs. Nor does it explain how the program's activities (treatments) are thought to lead to outcomes; it lacks a discussion of intervening mechanisms (Chen, 1990). In the article's discussion of the program's theory of change, it is reported that the staff discussed their perceptions of a theory of change for the program, but this does not appear to have been included in the logic model. Similar to the six studies described by Birkmayer and Weiss (2000), Yampolskaya, et al.'s discussion moves in the direction of a more holistic approach, but it lacks a fully articulated program theory. The concept

mapping tool does appear to be a good way to clearly articulate treatments (i.e. program activities), which suggests it may hold promise for articulating outcomes, as the current study will explore.

Rosas (2005) also used concept mapping, and he notes that it has not frequently been used to articulate program theory, despite Leeuw's (2003) recent suggestion. Rosas presents the results of a concept mapping analysis of outcomes for a family support program, and he speculates about how those results could inform the development of a program theory. But no actual program theory is articulated as part of Rosas's study. He writes: "It is important to recognize that what emerged from this process [i.e. concept mapping] was theorizing about outcomes, not a complete program theory. Thus, it would be inaccurate to claim the concept map represents a more comprehensive program theory without further work" (2005, p. 399). Rosas suggests that a more complete program theory could be built if concept mapping were also used to depict the activities and processes of the program in addition to the outcomes.

Two other recent examples exist which also begin to connect concept mapping and program theory. Sridharan, Campbell and Zinzow (2006) provide an example of concept mapping being used to supplement existing program theory for a juvenile violence reduction program. They used concept mapping to elicit anticipated program outcomes. Their rating process was used to rate not only the relevance of the outcomes but also the expected time to impact for each outcome. In other words, participants judged how much time would elapse between program implementation (treatment) and observation of specified outcomes. The results are presented as line graphs and then translated into a "rather crude but nevertheless useful program logic" (2006, p. 156),

which could be integrated into existing program theory. As another example, Trochim, Marcus, Masse, Moser and Weld (2008) use concept mapping to develop a logic model of outcomes for the evaluation of a large tobacco research initiative sponsored by the National Cancer Institute. Outcomes were modeled using the concept mapping process and "…an outcome logic model was developed by arranging the clusters of the concept map in the expected temporal order" (Trochim, et al., 2008, p. 11). Arranging outcomes in temporal order is not an example of a complete program theory, but it could be a step along the way. For example, from Dubin's (1978) method, it could be considered an enumeration of laws of interaction among a theory's units. Despite the lack of a complete theory, these two recent studies provide examples of concept mapping's use to examine programmatic outcomes. The current study extends such ideas.

Summary

The preceding review provides necessary background to understand theory-building methodology, theory-based evaluation and concept mapping. As discussed, theory-based evaluation still has many challenges to overcome in order to realize its full potential. Furthermore, the concept mapping tool is increasingly used in a variety of research contexts. Leeuw (2003, p. 106) notes the possibility of using concept mapping as an "elicitation methodology" for articulating program theory in support of theory-based evaluation. However, research has only begun to explore this idea. There appear to be no examples in the literature of concept mapping's use for articulation of a complete program theory. Nor are there any attempts to evaluate in a disciplined fashion how well concept mapping works for theory articulation. This study will explore whether concept mapping is an appropriate tool for conceptualizing and articulating the outcomes

domain of a program theory. It will access a broader literature to use ideas from Dubin (1978) and Whetten (1989) – in addition to those of Chen (1990) – to evaluate how well the concept mapping tool works for building theory.

Chapter 3 – Methods

This chapter introduces Metro Organizations for People (MOP), a community organization in Denver, Colorado. MOP's community organizing intervention provides the context for exploring whether or not concept mapping can conceptualize an outcomes program theory domain for a complex social intervention. After introducing MOP, details of the concept mapping method are provided, including a brief reflection on reliability and validity. Finally, the specific data collection process used at MOP is discussed, as well as the processes and analyses used to answer the research questions posed in Chapter 1.

Metro Organizations for People – Program Description

As mentioned previously, the context for applying concept mapping to the problem of program theory development is a grassroots community organization in Denver, Colorado, called Metro Organizations for People (MOP). Its website describes MOP as follows: "MOP is comprised of 35 member organizations including 11 schools, 18 congregations and 4 youth and neighborhood groups" (Metro Organizations for People, n.d.a). In order to understand the character of MOP's work, one must also understand the nature of the national network to which MOP belongs. That network is called the PICO National Network: People Improving Communities through Organizing. A description of PICO follows:

> PICO is a national network of faith-based community organizations working to create innovative solutions to problems facing urban, suburban and rural communities…. With more than 1,000 member institutions representing one million families in 150 cities and 17 states, as well as a growing international

> effort, PICO is one of the largest community-based efforts in the United States. (PICO National Network, n.d.)

As the description of PICO suggests, its member institutions, like MOP, are interested in solving community problems. Significant here is the fact that problems are described as plural; PICO organizations are not focused on a single issue or problem. MOP currently has committees which are working on a number of different issues. These include access to healthcare, local and state education reforms, college access, citizenship issues and neighborhood safety (Metro Organizations for People, n.d.d).

MOP conducts its work by practicing a style of community organizing promulgated by the PICO National Network. This organizing process begins in the context of one of MOP's member organizations, such as a congregation or school. Each member organization has a local organizing committee (LOC) composed of several leaders. Participants in a PICO affiliate's organizing process are referred to as "leaders". These leaders conduct a series of face-to-face visits with community members known as *1-1* (pronounced "one-to-one") visits (Snyder, n.d.). The purpose of these visits is to "build relationship[s], listen for concerns and invite participation" from other members of the church or school community (Snyder, n.d., p.1). Such visits begin to raise awareness among members of a congregation or school community and among the LOC leaders about the types of problems which affect that community. As Brazilian educator Paulo Freire observes, increased consciousness regarding an oppressive community problem is a prerequisite for community members to take action in solving that problem (2000).

After making the one-to-one visits, leaders complete a research process by meeting with public officials. Research meetings allow leaders to accomplish three things: (1) identify possible solutions to issues or concerns surfaced in the one-to-one

meetings, (2) "test their own public skills" and (3) "develop relationships with public officials" (Snyder, n.d., p. 1). Research is followed by action. Such action occurs in the form of "large community meetings where [the] community comes together to display political power and win commitments from public officials" (Snyder, n.d., p. 1). Finally, LOC leaders reflect on the entire process of visiting, researching and acting before beginning the process anew (Snyder, n.d., p. 1).

Sample

The sample for this study is a group of leaders from Metro Organizations for People who live in Denver or the surrounding area. The group included 20-30 people who are experienced leaders in the organization. Such a sample can be considered a "purposive" sample of "typical instances" in the sense described by Shadish, Cook and Campbell (2002, p. 23). That is, the study aims to include people who have had a typical experience of the organizing intervention. The choice of such a sample is guided by Trochim's suggestion that the people who are best suited to conceptualizing a program's theory are those who are closest to the intervention process (Trochim, 1998).

Details of Concept Mapping Tool

Concept mapping was first presented as a cohesive process in a 1989 article by William Trochim (1989). In the article, Trochim suggests that concept mapping is an example of a structured conceptualization process. Such processes, through a series of procedurally-oriented steps, produce a visual representation of some concept or idea (Trochim & Linton, 1986). Concept mapping is presented as a conceptualization process which is particularly helpful for program planning and evaluation. Trochim (1989, p. 1) describes concept mapping in six steps: (a) preparation, which includes selection of

participants and determination of the brainstorming and rating focus statements, (b) generation of [response] statements via brainstorming, (c) structuring of statements via sorting and rating, (d) representation of statements by computing a concept map, (e) interpretation of maps and (f) utilization of maps. Figure 2 provides an overview of the six steps. Each step will be explained in more detail below.

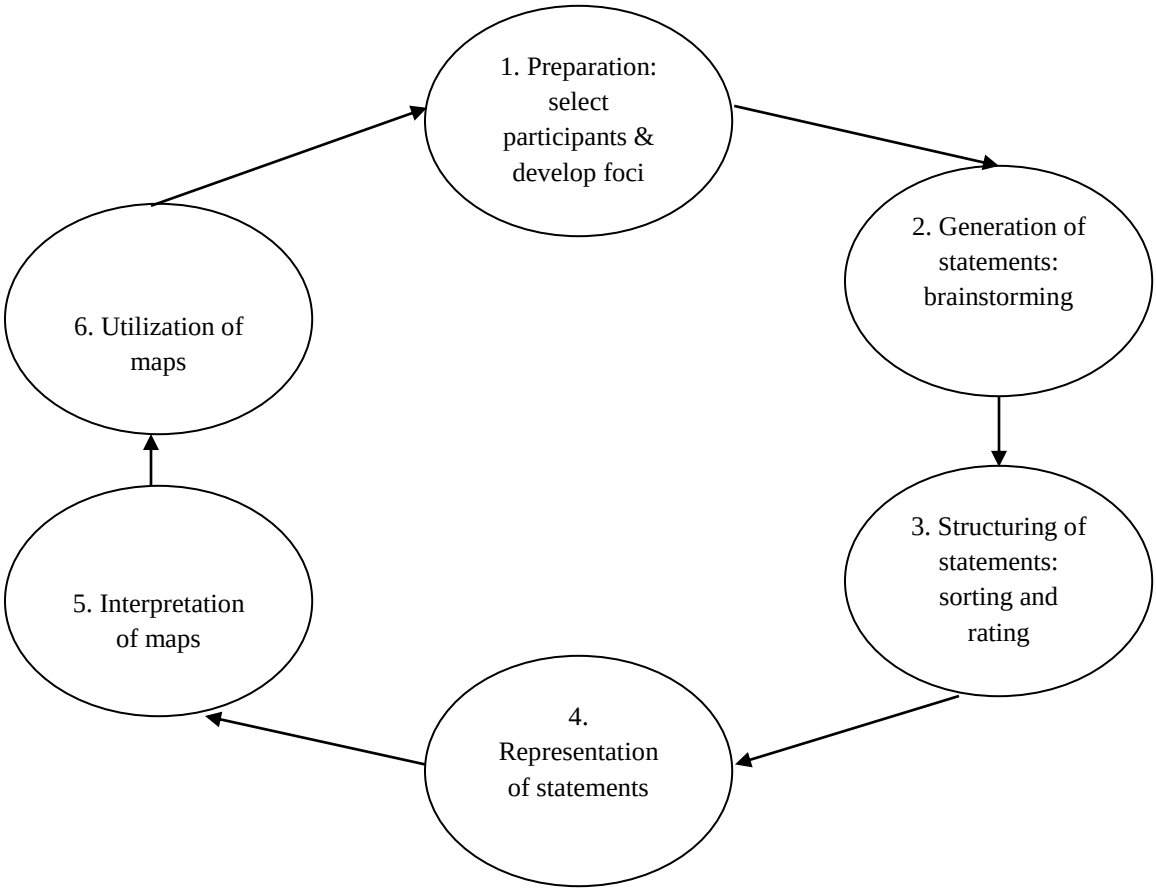


Figure 2. Six steps of the concept mapping process (Kane & Trochim, 2007; Trochim, 1989).

The preparation step includes deciding what group of people will produce the concept map. Trochim advocates a diverse group of participants (between 10 and 20), though he notes that concept mapping can be used with larger groups, too. The group can be selected via either purposive sampling (Shadish, et al., 2002) or random sampling (Trochim, 1989). The participant group is asked to respond to two focus statements; the brainstorming focus will help them generate response statements and the rating focus will help them rate the importance of each response statement. Trochim notes the importance of carefully wording the brainstorming focus statement so as to avoid a focus which actually contains two questions; the brainstorming focus must ask participants about one concept only. It is worth noting here that Trochim uses the word *statement* to describe two different pieces of the concept mapping process. For clarity, the reader should keep in mind that the two *focus statements* (brainstorming and rating) are developed by the researcher and are presented to the participant group at the beginning of a brainstorming or rating exercise. *Response statements* are statements that participants make in response to the brainstorming focus statement. Response statements constitute participant-generated data for the project.

The second step is generation of response statements (Trochim, 1989). This step is most often accomplished via a brainstorming exercise (W. Dunn, 1981; Osborn, 1948) where participants respond to the brainstorming focus statement. At this initial stage, there is no editing of response statements; all responses are recorded. However, participants are encouraged to ask for clarification when they do not understand the meaning of another participant's response. Later on, the set of response statements may be narrowed down by taking a random sample (see Linton, 1989) or by using a key-

words-in-context approach (Krippendorf, 1980, 2004; Stone, Dunphy, Smith, & Ogilvie, 1966) from content analysis. Narrowing the response set makes analysis more manageable. Kane and Trochim recommend not exceeding 100 statements so as to avoid "...excessive time for data input, unnecessary redundancy of the content, and a loss of group energy" (Kane & Trochim, 2007, p. 59).

After generating response statements, the third step in the concept mapping process is structuring of statements. Each response statement is recorded on a separate (numbered) card and participants engage in a card sort activity (Rosenberg & Kim, 1975). Participants are asked to sort the response cards "in a way that makes sense to you" (Trochim, 1989, p. 5). Participants may sort cards into any number of groups. However, three rules govern the sorting: (a) participants may not place all cards in one group, (b) participants may not place each card in its own group (so that the number of groups equals the number of cards) and (c) participants may not place any card into more than one group. After sorting, each participant's groupings are converted into a square, binary matrix and the binary matrices are summed to form a group similarity matrix (Trochim, 1989). (Details of matrix construction are discussed below in the "Data Analysis" section of this chapter.) Finally, to complete the structuring step, each response statement is also rated by each participant according to the importance s/he attaches to that particular statement. Often a simple Likert-type scale (Likert, 1932) is used for rating.

The fourth concept mapping step is representation of the response statements as a concept map. First, the group similarity matrix is used as input data for a statistical process called multidimensional scaling (Kruskal & Wish, 1978). Multidimensional

scaling uses the group similarity matrix to place the response statements on a map in a way that reflects the relative conceptual proximity of the statements (Trochim, 1989). Statements which were often sorted together will be close on the map and statements which were infrequently sorted together will be far apart. Although multidimensional scaling can produce a map in many dimensions, Trochim recommends a two-dimensional map because it is easy to understand. After creating a map on which there is a point for each response statement, a cluster analysis (Kaufman & Rousseeuw, 1990) is performed using each point's coordinates as input (Trochim, 1989). Cluster analysis groups the map points (i.e. response statements) into any specified number of clusters using the relative distances between points (Johnson & Wichern, 2007). Trochim suggests examining several different cluster solutions until one is found which makes the best sense. Ideally, the final cluster solution is informed by input from participants. Trochim recommends using Ward's clustering algorithm (Johnson & Wichern, 2007) rather than some of the other available methods. After a cluster solution is obtained, then the ratings for each response statement can be used to calculate an average rating for each response and for each cluster. At the conclusion of this fourth step, Trochim notes that the concept mapping process has produced: (a) a point map where numbered points represent each response statement, (b) a cluster map showing how the points are grouped, (c) a point rating map that displays the average rating value for each response statement and (d) a cluster rating map which displays the average rating for each cluster.

The fifth step involves interpretation of the maps. For this step, the participants are reassembled and presented with the results of the concept mapping exercise. The goal of this step is for the group to arrive at a consensus regarding a name for each cluster

and perhaps also a decision about whether there are logical groups of clusters that might represent meaningful regions on the map. Trochim writes: "This final named cluster map constitutes…the basic result of the concept mapping process…. it is [now] useful…to engage the participants in a general discussion about what the map tells them…" (Trochim, 1989, p. 11). The group may also examine the cluster map with average ratings in an effort to determine if the ratings make sense. Figure 3 provides an example of a named cluster map presented in Trochim's article.

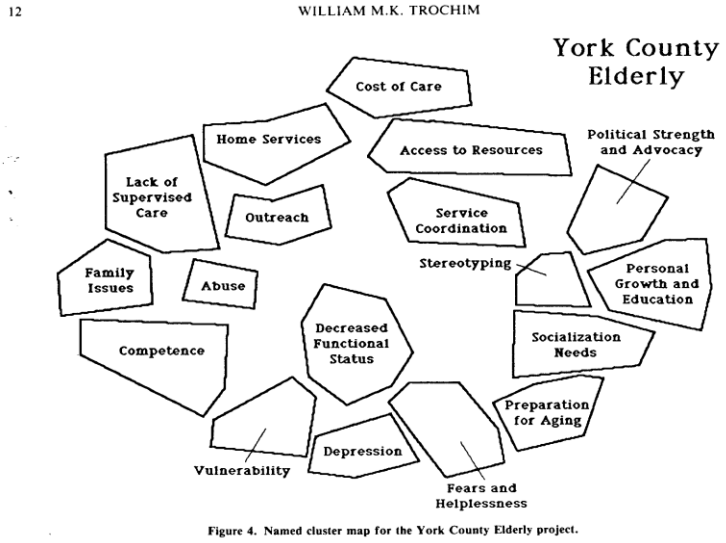


Figure 3. Sample named cluster map from Trochim (1989). Used with permission.

Finally, the sixth step consists of utilization of the maps. Trochim (1989) suggests a number of ways that the maps can support program planning and/or evaluation. Of particular interest to the current methodological study is the idea that the map can be understood as a representation of the program's *construct* and also that the map may be used as a guide to developing measures to evaluate the program's outcomes

(Trochim, 1989, p. 13). Such uses are consistent with ideas expressed in the literature on theory-based evaluation.

In 2007, Mary Kane and William Trochim published a comprehensive book about concept mapping (Kane & Trochim, 2007). Their text shows that the tool has evolved only slightly since the publication of Trochim's first article (1989). The six basic steps remain the same and the same output is still produced. However, Kane and Trochim have added two additional pieces of output, the pattern matching display and the go-zone graph. The pattern match can be a graphical comparison of cluster ratings for two different groups. For example, cluster ratings could be compared for two groups of participants or for concept maps developed at two different points in time (Kane & Trochim, 2007). The go-zone graph can depict the ratings of two different participant subgroups for each point in a single cluster. The graph is divided into quadrants and the name go-zone derives from the fact that points in the upper right quadrant of the graph represent "statements of a cluster that were rated above average" by both groups (Kane & Trochim, 2007, p. 22). Depending on the context of a planning, evaluation or research project, the information in the pattern match or the go-zone graph may be helpful.

Subsequent chapters of Kane and Trochim (2007) are devoted to each step of the concept mapping process. They make suggestions and updates which go beyond the original article (Trochim, 1989). For the preparation step (step 1), new suggestions include running a pilot test to see if the brainstorming focus statement generates the kind of response statements expected by the researcher. For generating response statements (step 2), a new web-based format is presented. More suggestions are also provided for how to reduce or edit the set of response statements. With regard to structuring response

statements (step 3), the text contains several templates to capture sorting and rating data and it presents a remote, web-based structuring process. Also, alternative structuring methods (besides the card sort) are discussed.

The fourth step of the process has been renamed a "concept mapping analysis" (Kane & Trochim, 2007, p. 12) instead of a "representation of statements" (Trochim, 1989, p. 7). Key components remain the same, however, including use of a similarity matrix, multidimensional scaling and cluster analysis. One addition is a brief discussion of goodness-of-fit measures for multidimensional scaling. Two other suggestions for analysis include determination of anchoring versus bridging statements on the map and use of the INDSCAL algorithm (Kruskal & Wish, 1978) for a simultaneous scaling of participants and statements. The former would be difficult to implement in this study, however, because the calculations for anchoring and bridging indices are described as proprietary to Kane and Trochim's software package (Concept Systems Incorporated, 2005).

Kane and Trochim (2007) provide practical suggestions for conducting step 5 (interpretation). They then present separate chapters for the sixth step: using the maps. One chapter deals with using concept maps for planning and the other with using maps for evaluation. In the chapter on evaluation, Kane and Trochim discuss the use of concept mapping to build a logic model, to develop questions for an evaluation and to develop measures and scales. All of these ideas are consistent with ideas presented in the literature about theory-based evaluation, and they suggest a connection to more general processes of theory-building, such as Dubin (1978). Kane and Trochim cite Rosas (2005) as an example of how concept mapping "can be used to develop program theory" (Kane

& Trochim, 2007, p. 166). However, as noted in the literature review in Chapter 2, a closer look at Rosas's study shows that he does not use concept mapping to articulate a complete program theory.

Reliability of Concept Mapping

Trochim (1993) presents several suggestions for assessing the reliability of concept mapping in an American Evaluation Association conference paper. He notes that traditional conceptions of reliability, which consider the consistency of scores for a particular instrument and a specified population (Morgan, Gliner, & Harmon, 2006), may not be appropriate for concept mapping. He suggests measures for concept mapping which shift the focus from consistency of *items measuring constructs* to consistency of *persons articulating concepts*.

Trochim (1993) makes several suggestions for measuring reliability. These are: (a) test-retest reliability, which correlates group similarity matrices or map distances between point pairs from two sorts of the same response statements at two different times, (b) split-half reliability, which correlates group similarity matrices or map distances between point pairs from a random division of participants into two subgroups, (c) average split-half reliability, which averages the split-half reliabilities for all possible splits of the data into two subgroups, (d) average individual-to-total or individual-to-map reliability (cf. average item-total reliability), which averages correlations between each individual participant's matrix and the group similarity matrix or between each individual matrix and the distance pairs on the map and (e) average individual-to-individual sort reliability or average rating-to-rating reliability (cf. average item-item reliability), which correlates each possible pair of matrices or rating score vectors from among all

participants and then takes the average of the correlations. Trochim concludes that (a) and (c) above are generally not practical. In the case of (a), it is not often possible to reassemble a group of participants at a later date for re-sorting the statements. In the case of (c), Trochim believes that average split-half reliability is not computationally feasible. Removing the test-retest option and reorganizing Trochim's list to give each calculation its own letter, yields six possible reliability computations. These six are denoted in the following list with primes (') to distinguish them from Trochim's original list: (a') split-half matrix reliability, (b') split-half map reliability, (c') average individual-to-total reliability, (d') average individual-to-map reliability, (e') average individual-to-individual sort reliability, and (f') average rating-to-rating reliability (Trochim, 1993).

Trochim also suggests that the reliability calculations discussed above should be adjusted by the Spearman-Brown prophecy formula (Brown, 1910; Spearman, 1910) to account for the fact that reliability correlations are based on only part of the participant sample (Nunnally, 1978). Trochim's suggestion to use Spearman-Brown directly extends the use of this formula from traditional reliability theory into reliability calculations for concept mapping. From the literature, however, it appears that the correct way to use the Spearman-Brown formula for *concept mapping* is somewhat unclear (see Bedi, 2006; Cacy, 1996). Trochim writes:

> …since we know that reliability is affected by the number of items on a test (or persons in a concept mapping project), these correlations based on only part of the participant sample do not accurately reflect the correlational value we would expect for the entire participant sample. This is traditionally corrected for in reliability estimation by applying the Spearman-Brown Prophecy Formula. (1993, "Estimates of Reliability", para. 10)

Trochim cites Nunnally (1978, p. 211) for background on the derivation and application of the Spearman-Brown formula. Nunnally's discussion is based on a domain theory of

reliability. The derivation of formulas is complex and they depend on sampling theory related to both the number of people studied and a hypothetical domain of test items. In order to be certain that the extension of domain theory and the Spearman-Brown formula from traditional reliability theory to concept mapping is correct, additional research is certainly needed.

Now, consider four studies which calculate at least some of Trochim's (1993) reliability measures. First, Trochim, Cook, and Setze (1994) use split-half reliability and individual-to-individual sort reliability in their study of program staff perceptions of an employment program for individuals with mental illness. Split-half matrix reliability (a') and split-half map reliability (b') are .79 *(df*= 4,559, $p < .001$) and .56 *(df*= 4,559, $p <$ .001), respectively, and are adjusted for Spearman-Brown. The average individual-to-individual sort reliability (e') was calculated using a contingency coefficient (McNemar, 1969) for each of the possible pairs which could be formed among participants. The average contingency (reliability) coefficient was .85. Second, Cacy (1996) provides all six of Trochim's reliability measures in his study on the nature of a practice-based research network with two different groups of physicians, one that practiced in the community and one composed of academic faculty. Average individual-to-individual reliability (e') was .61 and average individual-to-total reliability (c') was .81. Split-half matrix reliability (a') was .60 and split-half map reliability (b') was .39. Finally, average individual-to-map reliability (d') was .81 and average rating-to-rating reliability (f') was .88. All of these measures were adjusted for the Spearman-Brown correction. Third, Rosas (2005) finds that the split-half matrix reliability (a') is .83 (df =4,464, $p < .001$) and the average individual-to-total reliability (c') is .94 (df = 4,464, $p < .001$). Both are

adjusted for Spearman-Brown. Finally, Bedi (2006) calculates reliabilities for a study of alliance formation between counselors and their clients. He uses three of Trochim's (1993) calculations. The average individual-to-individual sort reliability (e') was .17 ($p <$ .05). The average individual-to-total correlation (c') was .45, with no *p*-value stated. Finally, the split-half matrix reliabilities (a') were .76 ($p < .001$) and .74 ($p < .001$) for different splits of the data without using the Spearman-Brown correction.

The empirical results from these four studies suggest that concept mapping produces reliable output (i.e. good consistency of persons articulating concepts); most coefficients are .60 or higher. Clearly, however, concept mapping would benefit from a more in-depth analysis of how transferable are the traditional item-based reliability assumptions to a participant-based reliability framework. This is especially true for adjustments like the Spearman-Brown prophecy formula (Brown, 1910; Spearman, 1910).

Validity of Concept Mapping

Next, consider the validity of knowledge produced by concept mapping. As a starting point, recall Shadish, Cook and Campell's (2002, p. 34) statements that (a) *validity* refers to "…the approximate truth[1] of an inference" (or knowledge claim or proposition) and (b) "…validity is a property of inferences. It is *not* a property of designs or methods…." Thus, when discussing validity and concept mapping, the question is not: "Is concept mapping a valid method?" Rather, the relevant question is: "Does concept mapping produce valid knowledge in a specified situation?" Trochim's suggestion that

[1] Shadish, Cook and Campbell note that the definition of "truth" varies among different schools of philosophy. In applied research, it would also depend on the researcher's ontological and epistemological assumptions.

concept mapping be understood as a process of persons articulating concepts will also be taken as the framework for validity discussions. Thus one can ask: "Does concept mapping enable a group of people to produce a valid (i.e. "truthful") articulation of the concept under study?"

A few authors have discussed how to assess the validity of results articulated by concept mapping, but these discussions are not as developed as the discussions on reliability. The earliest discussion relevant to the validity of articulated concepts is Dumont's (1989) study of multidimensional scaling. Dumont asked whether maps formed by multidimensional scaling were a valid representation of a participant's conceptualization of "factors contributing to living in the community and reinstitutionalization from the perspective of persons psychiatrically institutionalized two or more times" (Dumont, 1989, p. 82). Although she does not call the method concept mapping, she used multidimensional scaling (MDS) and cluster analysis in a similar fashion to Trochim's (1989) sense of concept mapping. One large difference, however, is that Dumont produced an MDS-computed map for each individual participant while concept mapping normally involves the aggregation of data matrices for multiple participants before the application of MDS. Any conclusions about the validity of concept maps which are drawn from Dumont's study must account for this basic difference.

After producing an MDS map for each of five participants, Dumont asked each person to hand-place his or her own clustered statements on a map to form a basis for comparison to the MDS maps. Euclidean distances were calculated between cluster centers on both the MDS maps and the hand-placed maps. These distance pairs were

correlated for individual clusters and for regions of clusters. (The author does not give enough detail to completely understand the regional cluster correlation analysis.) Dumont's results show that regional clusters display very low or even negative correlations, while individual clusters show higher correlations. Correlation values vary significantly among the five participants. Thus, the results of Dumont's study are quite mixed in terms of whether MDS can provide a valid representation of an individual's hand-placed conceptualization of statements. And one must also ask whether even the hand-drawn "theoretical" (Dumont, 1989, p. 81) representations accurately capture an individual's conceptualization; it is not clear that these theoretical maps should be the *a priori* basis for comparison. In the end, this study provides little evidence either for or against the validity of MDS representations of what an individual really thinks. And it provides no evidence for the larger question of the validity of concept maps when used with a group instead of with individuals.

Cacy (1996) explored the validity of concept maps in a slightly different fashion than Dumont (1989). His study followed Trochim's (1989) concept mapping process more closely than Dumont's. Cacy produced three concept maps relating to the nature of a "practice based research network" (1996, p. 60): one for each individual physician group and one for the combined group. He then asked each participant to choose the map that "makes the most sense" (1996, p. 95). The results showed that the faculty group more consistently chose the community practitioners' map than their own map. The community practitioners did not consistently pick one map. In the study, Cacy understands validity to mean whether or not the maps are "…in any sense real to the participants" (1996, p. xi). Overall, he concludes that his study provides "…no

compelling evidence for the validity of the concept maps generated during the study" (1996, p. 108).

A useful frame for thinking about the validity of concepts articulated via concept mapping is that used by Jackson and Trochim (2002), who take a perspective from content analysis. They are positive about concept mapping's potential validity in representing a group's perceptions. They note:

> The main strength that concept mapping offers to validity is that by using multi-dimensional scaling and cluster analysis to represent the similarity judgments of multiple coders, it allows meaning and relationships to emerge by aggregating the 'biases' or 'constructions' of many. Instead of arbitrary bias and potentially forcing values of the investigator with a priori categories or semantic encoding choices, sorting concepts allows for a web of concept relationships to be represented by sorters immersed in the context of their own social reality. (Jackson & Trochim, 2002, p. 330)

Jackson and Trochim suggest here that collective conceptualizations from concept mapping are potentially more valid than are results from other methods which rely more on the researcher's role or interpretations. However, they agree with Krippendorf when they write that because concept mapping deals with social constructions, "…there is really no way to establish a standard by which to judge the degree of error" in the expression of participants' perceptions (2002, p. 330; Krippendorf, 1980). Fortunately, the concept mapping method has a validity check built into its process, namely the review and interpretation of maps by the participant group. If the maps do not make sense to the participant group, then the researcher can conclude that the articulation of concepts lacks validity and s/he must find a better tool to use. In the end, discussions about both reliability and validity are best summarized by Bedi and Alexander: "No single, infallible estimate of either reliability or validity for this multistep complex process is available (Bedi & Alexander, 2009, pp. 83-84; Trochim, et al., 1994).

Data Collection

The method for applying concept mapping in this study generally followed the process outlined by Kane and Trochim in their recent text *Concept Mapping for Planning and Evaluation* (2007) and by Trochim in his original concept mapping article (Trochim, 1989). It was also guided by the experiences of a small pilot study conducted with 11 MOP board members in November, 2009. The concept mapping process began with a brainstorming activity which encouraged participants to think about any outcomes which they believe have occurred since they got involved with MOP. Participants were encouraged to think broadly about outcomes that have occurred in their city, in their neighborhood, in their children's schools, in their church communities, in their families or in themselves. Participants were asked to respond to this prompt: "Think about yourself, your family, your child's school, your church and your neighborhood. When MOP does community organizing, this is what happens: _____." Participants did not need to come to any consensus on whether or not a specific response to this prompt was an accurate reflection of outcomes. Rather, all responses were recorded. Verbal instructions for brainstorming and all responses spoken in English during the brainstorming session were translated into Spanish for participants who either did not speak English or who were more comfortable speaking Spanish. All questions and responses made in Spanish were translated for English-speakers to understand. This Spanish-English translation process was familiar to MOP leaders, as they frequently use a translator and translation equipment for their own meetings.

After the brainstorming exercise, a set of cards was produced for each participant with one brainstormed response printed on each card (Kane & Trochim, 2007). The

cards were also numbered so as to facilitate recording of results. After the brainstorming meeting, packages were assembled for each participant that contained a set of cards (one for each statement) and also instructions and recording sheets for the sorting and rating activities. Each participant received a package by mail either at their home or at the MOP office, according to their preference. Statements and instructions were translated so that a Spanish-language packet could be provided to any participant who requested it. Postage-paid envelopes were provided in the packet so that each participant could return the results of his or her sorting and rating exercises at no cost.

Sorting is the next activity in the concept mapping process (after brainstorming). The sorting instructions asked participants to sort the response cards in any way that makes sense to them (Bedi & Alexander, 2009; Carter, Enyedy, Goodyear, Arcinue, & Puri, 2009; Kane & Trochim, 2007). Participants then recorded their card groupings on a sheet provided in the package. Participants could sort cards into any number of groups that they chose. They were asked to follow the three sorting rules mentioned above (i.e. participants may not place all cards in one group, they may not place each card in its own group and they may not place any single card into more than one group, Kane & Trochim, 2007).

The next step in the concept mapping process is usually rating, where each participant is asked to rate each response statement on an ordinal scale (Trochim, et al., 1994) to express how important the outcome is to him or her. The goal of the rating exercise is to calculate an importance rating for each cluster of responses on the final map. However, in prior studies using the concept mapping technique, it appears that cluster importance ratings are often very close to one another and they exhibit little

variability between clusters (e.g. Bedi, 2006; Paulson & Worth, 2002). This low-variability phenomenon was also observed in the pilot study with MOP. In the pilot data, the majority of response statements had an average rating between 4 and 5 (i.e. between "very important" and "extremely important"). No response had an average rating below 3 (i.e. "important"). Thus, instead of rating each response statement according to perceived importance, current study participants rated each response (i.e. each MOP outcome) as "short-term" or "long-term". The prompt for this timeframe rating exercise was: "Please tell me whether each statement listed below occurred in the **short-term** (took less than six months of organizing) or in the **long-term** (took more than six months of organizing)." There were two choices: "short-term" and "long-term", which were then numerically coded with values of 0 and 1, respectively. A "timeframe" rating could be calculated by taking the proportion of responses within a cluster which were marked "long-term" by participants. The goal of the timeframe rating was to further understand how comprehensive are the potential outcomes units generated by the concept mapping. The literature on concept mapping suggests only one study which has modified the rating step of the concept mapping process. In Sridharan, et al. (2006), participants were asked to rate responses on the basis of expected time-to-occurrence instead of only on a criterion of importance.

The list of sorted and rated statements was assembled into a two-dimensional concept map using the R statistical software package (R Development Core Team, 2011). The structure of outcomes was represented by every outcome's relative location with respect to every other outcome in the conceptual space depicted on the map. Outcome structure was further represented by the clusters of outcomes on the concept map. The

timeframe of various outcomes was represented by an average timeframe rating for each cluster, which was derived from the rating data.

Data Analysis

In order to perform multidimensional scaling (Kruskal & Wish, 1978) and cluster analysis (Johnson & Wichern, 2007; Kaufman & Rousseeuw, 1990) to build the map, it was necessary to first perform a significant reformatting of the response data. Reformatting converted the data from its collected format (a grouped list of statements for each participant) into a matrix format suitable for analysis. The reformatting took place as follows. First, each response statement from the brainstorming session was numbered from 1 to S, where S indicates the total number of statements. For purposes of the following discussion, let s_i represent the i^{th} brainstormed response and s_j represent the j^{th} brainstormed response. Each participant was numbered from 1 to P, where P indicates the total number of participants and p_i represents the i^{th} individual participant. Each participant's sort groups (as recorded on the sheets in their package) were transformed into a symmetric binary similarity matrix (BSM) for that participant (Bedi, 2006). There was one binary similarity matrix for each of the P participants. Each BSM was square and symmetric with dimension S x S. Each row and column in the BSM represented one response statement from the brainstorming session. Figure 4 below, along with the following discussion, provides an example of a BSM.

	s_1	s_2	s_3	-------	-------	-------	s_{78}	s_{79}	s_{80}
s_1	1	0	1	-------	-------	-------	0	0	0
s_2	0	1	0	-------	-------	-------	0	1	0
s_3	1	0	1	-------	-------	-------	0	0	0
⋮	⋮	⋮	⋮	⋱	⋱	⋱	⋮	⋮	⋮
s_{78}	0	0	0	-------	-------	-------	1	0	0
s_{79}	0	1	0	-------	-------	-------	0	1	0
s_{80}	0	0	0	-------	-------	-------	0	0	1

Figure 4. Sample binary similarity matrix (BSM) for one participant with 80 sorting statements. Statements 1 and 3 and statements 2 and 79 are sorted together.

Assume there are 80 brainstormed statements (S=80). Then the binary similarity matrix for the p_i^{th} participant (see Figure 4) has 80 rows and 80 columns. Both column 1 and row 1 represent the first of the 80 brainstormed statements (s_1). Row 2 and column 2 represent the second of the 80 brainstormed statements (s_2), and so on. An individual cell in the BSM represents an association between the two responses whose row and column determine the cell. In participant p_i's binary similarity matrix, a 1 appears in the cell at

the intersection of row s_i and column s_j if participant p_i sorted responses s_i and s_j into the same group. Similarly, there is a 0 in the cell at the intersection of row s_i and column s_j if participant p_i did not sort responses s_i and s_j into the same group. The cell at the intersection of row s_j and column s_i contains the same value (either 0 or 1) as the cell at the intersection of row s_i and column s_j. If, as shown in Figure 4, the p_i^{th} participant sorted responses 1 and 3 into the same group, then the cell at row 1 and column 3 and the cell at row 3 and column 1 in the BSM both contain the value 1. Since the p_i^{th} participant did not sort statements 1 and 2 together, the relevant cells are zeros. All values along the diagonal of a BSM are 1 because it is assumed that each statement is sorted with itself.

After a binary similarity matrix was created for each of the P participants, these BSMs were then summed to form a GSM or group similarity matrix (Bedi, 2006). Values of the cells in the group similarity matrix represent how many times each pair of statements was sorted together by participants. The values in the GSM can range from 0 (indicating that a pair of statements was never sorted together by any participant) to P (indicating that a pair of statements was sorted together by every one of the P participants). As Bedi notes, "the value in this latter matrix [the GSM] for any pair of statements indicates how many participants placed that pair of statements together in a pile regardless of what other statements were included or excluded from that pile" (2006, p. 28). Finally, the GSM was transformed into a group dissimilarity matrix (GDM). This was accomplished via a monotone decreasing transformation (Buja, et al., 2008) so that pairs of statements which were not often sorted together (and thus had low similarity in the GSM) now had a high dissimilarity value in the GDM. The transformation subtracts each GSM value from P to obtain a dissimilarity value for each cell in the GDM

(Kaufman & Rousseeuw, 1990). Values in the GDM still ranged between 0 and P, but now a high value of P indicates that a pair of statements was *not* often sorted together.

The group dissimilarity matrix formed the basis for a non-metric multidimensional scaling (MDS) analysis (Bedi, 2006; Carter, et al., 2009; Trochim, 1989). Details on the nature of this analysis are contained in Kruskal and Wish (1978). Essentially, however, MDS takes the information in the group dissimilarity matrix and expresses it as a distance between each pair of points. A high dissimilarity value corresponds to a large distance between points. These distances are used to build a map showing the relationship of each one of the S response statements to each one of the remaining (S-1) statements. An MDS analysis can provide a map solution in any number of dimensions specified by the researcher. Usually, a two-dimensional solution is used with concept mapping due to its ease of representation (Trochim, 1989). It is often difficult to visualize a solution in three dimensions and it is particularly difficult in four or more dimensions. This study follows the convention of using a two-dimensional solution.

After the S response statements were put onto a two-dimensional map, a second statistical technique, cluster analysis, was then performed to group mapped response statements into clusters (Trochim, 1989). Clusters are mathematically-based groupings of the statements. Clusters are also closely related to how often each pair of statements was sorted together by participants in the sorting activity. As with multi-dimensional scaling, there are multiple potential cluster solutions. Johnson and Wichern (2007, p. 695), in discussing hierarchical clustering methods, suggest trying multiple methods and, "…if the outcomes from the several methods are (roughly) consistent with one another,

perhaps a case for 'natural' groupings can be advanced". Kaufman and Rousseeuw also suggest using multiple clustering methods (Kaufman & Rousseeuw, 1990, p. 37). In particular, they value partitioning methods over hierarchical methods because partitioning methods are better designed to find a "best" set of cluster solutions (Kaufman & Rousseeuw, 1990, p. 44).

For this study, four types of clustering discussed by Kaufman and Rousseeuw (1990) were used: partitioning around medoids (PAM), fuzzy analysis (FANNY), agglomerative nesting (AGNES) and divisive analysis (DIANA). PAM, as the name suggests, is a partitioning method. Partitioning methods require specification by the user of *k*, the number of clusters desired. Then, "…the algorithm tries to find a 'good' partition in the sense that objects of the same cluster should be close or related to each other, whereas objects of different clusters should be far apart or very different" (Kaufman & Rousseeuw, 1990, p. 39). PAM chooses *k* representative objects called medoids and then builds clusters by assigning each object to the nearest medoid. FANNY is also a partitioning method. It does not definitively assign each object to one cluster, but rather assigns each object to every cluster with some percentage likelihood. For example, "…instead of saying 'object *a* belongs to cluster 1', FANNY can say that 'object *a* belongs…90% to cluster 1,…5% to cluster 2, and…5% to cluster 3,' meaning that *a* is probably to be assigned to cluster 1, but that there is still a glimpse of doubt in favor of clusters 2 and 3" (Kaufman & Rousseeuw, 1990, p. 42). A "hard" clustering can be obtained using FANNY by assigning each object to the cluster with the largest percentage likelihood (p. 44).

AGNES is a hierarchical clustering method which begins with every object in its own group and joins the objects together. Once a pair is joined, it cannot be split again. According to Kaufman and Rousseeuw (1990), this inflexibility is a limitation of hierarchical methods. Clusters of any size can be formed using AGNES. Finally, DIANA is also a hierarchical method. However, DIANA works in the opposite direction from AGNES. It starts with all objects in one group and then breaks the large group into any number of smaller clusters. All of the cluster analysis for this study was conducted using the 'cluster' package (Maechler, 2011) in the R statistical software environment (R Development Core Team, 2011). The 'cluster' package has fully implemented the four types of clustering just described, according to the algorithms of Kaufman and Rousseeuw.

The data analysis just described constitutes the fourth step of the concept mapping process, known as the "representation of statements" (Trochim, 1989, p. 7) or the "concept mapping analysis" (Kane & Trochim, 2007, p. 12). In the current study, step 4 yielded a point map, with one point for each response statement. It also yielded multiple cluster maps on which points are grouped into clusters and the cluster groups are indicated with ovals super-imposed on the map. Finally, average timeframe ratings (short-term vs. long-term) for each cluster were also superimposed on the maps. Following the concept mapping analysis, the final steps of the concept mapping process are interpretation and utilization. These final steps assessed the maps' value for articulating units which can be incorporated into a program theory of MOP's community organizing intervention.

Outcomes Domain Conceptualization

After results from all of the statistical analyses were obtained (including multiple possible clustering solutions), information from the maps and the associated analyses were presented to a group of participants during a follow-up interpretation session. The aims of this session were to understand what meaning participants would attach to the map and to lay the groundwork for answering the research questions about how well concept mapping produces units (Dubin, 1978) which can be used to build program theory. All of the participants' observations from the interpretation session were digitally recorded so that this information was available for consideration when answering the research questions.

The first two research questions ask whether potential outcomes units (as presented on the concept maps) conform with distinctions set out by Dubin (1978, e.g. thing vs. property of thing, attribute or variable) and with Whetten's (1989) requirements of comprehensiveness and parsimony for the subject matter of a theory. The study hypothesizes that each cluster of brainstormed responses on MOP's concept map represents a potential unit of theory (per Dubin's definition). These units lie within Chen's (1990) outcomes domain. The naming of these potential units occurred as part of the interpretation step of the concept mapping process. Participants were asked to label or name the clusters from several cluster solutions. Participants chose a final cluster solution, based on which of the clustered maps resonated most clearly with their understanding of the outcomes from grassroots community organizing. To support this process, participants used both the point map and two cluster maps. Participants also had a printed list of the statements belonging to each cluster. Materials were available in

English and Spanish to facilitate full participation. To answer the first research question, each participant-named cluster was later compared to Dubin's criteria in order to evaluate logically whether the named clusters could be considered units according to Dubin's definition (see Chapter 5). If the named clusters did not meet the criteria for units, then concept mapping would have produced only a conceptual framework for outcomes or perhaps a grouped list of events – but not the actual material from which a theory can be built.

To answer the second research question, participants assessed Whetten's (1989) criteria of comprehensiveness and parsimony. They were asked (a) whether any group of outcomes from the organizing intervention is missing from the concept map and (b) whether any of the named clusters are redundant. These questions were also addressed to a MOP staff member present at the interpretation session, as he was able to provide an additional perspective on the comprehensiveness and parsimony of the mapped representation of the community organizing process.

The last research question concerns how valid are potential units of theory produced by concept mapping. As noted in the discussion above, there is no agreed-upon set of criteria for assessing the validity or trustworthiness of the collective conceptualization of a phenomenon produced by concept mapping. This study explores two potential sources of evidence for the validity of a concept map. The first source of evidence is participants' judgments about the credibility (Lincoln & Guba, 1985) of the final map. The second source of evidence is a consideration of the similarity of groupings (i.e. potential units of theory) produced by different clustering methods.

Similar groupings resulting from different methods would suggest that concept mapping is providing a valid representation of programmatic outcomes.

Conclusion

At the conclusion of the interpretation session, all the necessary data to answer the research questions was available. Chapter 4 includes a presentation of all the concept maps with average timeframe ratings and the results of naming clusters during the interpretation session. Chapter 5 discusses the meaning of these results. It connects the results to the research questions and establishes whether or not the concept mapping tool is indeed able to articulate units for program theory-building.

Chapter 4 – Results

Generation of Statements – Brainstorming

On February 9th, 2011 two brainstorming sessions were held at Metro Organizations for People's offices in Denver, Colorado. Two sessions enabled broader participation by MOP leaders. The first session ran from 11:30 AM – 1:30 PM and had 14 attendees. The second session ran from 6 PM – 8 PM and had 6 attendees. MOP staff noted that other leaders had committed to come, but for various reasons, several cancelled at the last minute. MOP (as is the organization's custom) served a meal for participants at both sessions. Due to people arriving late and eating their meal, both sessions started about half an hour late. Both sessions opened with a review of the Colorado State University Internal Review Board (IRB) consent form and participants signed two copies, one for the project records and one to keep for themselves. Participants also completed a brief demographic survey, the results of which are discussed below.

The brainstorming prompt was as follows: "Think about yourself, your family, your child's school, your church and your neighborhood. When MOP does community organizing, this is what happens: _____." The prompt was printed on the meeting agenda, which was available in both English and Spanish. For the lunchtime session, there were several participants who were monolingual Spanish speakers. MOP employed a translator and the Spanish-speaking participants could hear all conversation translated into Spanish through a headset. When a participant spoke in Spanish, the translator then

spoke English, and all monolingual English-speakers could hear the translation via headset.

Both sessions produced very lively discussion in response to the prompt. While participants discussed their responses to the prompt, a staff member took notes. At the noon session, the note taker was MOP's Director of Operations. At the evening session, the note taker was MOP's Executive Director. After about 45 minutes of brainstorming, the groups took a short break. Then they reconvened briefly to review the list of statements they had created and to edit for clarification and removal of redundancies. At the end of the noon session, the group of 14 participants had produced 69 statements. At the end of the evening session, the group of 6 participants had produced 56 statements. The original statement lists for the noon and evening sessions are presented in Appendix A and Appendix B, respectively.

Although there were two separate brainstorming sessions for reasons of accommodating participant schedules, this study ultimately sought to create one concept map as a foundation for program theory. Thus, it was necessary to combine these two lists into one. Simply using all statements from both sessions would have resulted in 125 statements. Kane and Trochim (2007) suggest limiting the number of statements for sorting to about 100. Experience with analysis from the study's pilot project suggested that the list should be smaller in order to reduce sorting time and potential sparsity of the group similarity matrix (GSM). Therefore, the combined list was reduced to 89 statements after the brainstorming sessions were complete. Statements from both groups which reflected a similar theme or topic were combined so as to remove redundancy in the lists and to reduce the total number. For example, the statement "Meet people you

have never met before" (Appendix A, statement 4) and the statement "Helps me to get to know people with whom I would otherwise not have related to or known" (Appendix B, statement 9) were combined. The combined statement reads: "I get to know people I would otherwise not have known" (Appendix C, statement 5). Once the list had been reduced to 89 statements, there appeared to be no more statements which could be combined without some loss of information. The final list of 89 statements is included as Appendix C. This is the list that participants received in their packets for the sorting and rating activity.

Structuring of Statements – Sorting and Rating

The next step of the concept mapping process is the sorting and rating of the 89 statements. To accomplish this task, each participant received a packet that contained the following: instructions, pages for recording sort groups, a rating sheet and 89 cards (one statement per card) for sorting. The instruction sheet and the rating sheet are included as appendices D and E. English-language packets were mailed to 16 participants on February 11th, 2011 – two days after the brainstorming sessions. Four Spanish-language packets were mailed to participants who had requested such on February 16th – one week after the brainstorming sessions. Spanish packets were mailed later because all 89 statements had to be translated. Participants were asked to put their packets back in the mail by February 25th, 2011, using a postage-paid envelope provided in the packet.

Encouraging participants to return the packets proved to be a challenge, more so than the results of the pilot test had suggested. (For the pilot test, ten out of eleven packets were returned). By February 28th, only eight packets out of twenty had been returned. English-speaking participants (seven people) received a reminder call on

February 28th. On March 3rd, a bilingual office assistant at MOP followed up with the monolingual Spanish-speaking participants (three people). By March 11th, twelve packets had been returned and a second round of reminder phone calls was made.

The Director of Operations at MOP suggested recruiting additional participants for the sorting and rating activity, given the challenges of encouraging people to respond. She requested additional packets to distribute to advanced leaders at a weekend training session hosted by MOP. The material from the brainstorming session would be familiar to any experienced MOP leader, enabling them to sort the statements in a meaningful way. Sixteen additional English packets and six additional Spanish packets were provided. After the weekend training on March 19th and 20th, 2011, five packets were returned. In total, twenty-one sorted packets formed the basis for the multi-dimensional scaling and cluster analyses.

There was one packet for which the sorting results suggested that a participant had not understood the instructions for sorting and rating. In this particular packet, a blank card (numbered 90) had been accidentally included. The participant sorted the blank card into a group, in addition to the cards which contained statements. Also, this participant alternated the ratings of all 89 statements between short-term and long-term. Finally, the groups constructed by this participant tended to contain numeric values from the same decade (i.e. many 20's were in the same group, many 30's were in the same group, and so on). It seemed that this participant's data should perhaps be removed. However, after running the multidimensional scaling both with and without this participant's data, the goodness-of-fit metrics (described in more detail below) were very similar for both MDS runs. Thus, analysis continued with all twenty-one packets of data.

As the packets were returned, each one was transcribed into an MS Word document that was backed up on a local server. In a few participant packets, there were discrepancies between the groups which a participant listed on their sorting sheets and the way the cards were physically arranged into groups. In cases of a discrepancy, the physical arrangement of the cards determined which statement was in which group. After transcription, the R code to build a binary similarity matrix for each participant was tested. The R code contains two checks for accuracy. The first check counts the number of statements transcribed into each participant's groupings. Except in cases where a participant had a group containing only one statement, the number of transcribed statements was checked to see that it equaled 89. The second check in the R code counts all possible pairs within each group and sums the number of pairs across groups. Then, R calculates the length of an array of these pairs, which is twice the number of pairs. Twice the number of pairs should equal the sum of all entries in the binary similarity matrix. Each participant's binary similarity matrix was checked to be sure that this condition was true.

Participants' rating of each statement was also transcribed. Recall that each of the 89 statements was to be rated according to whether it was a "short-term" or "long-term" outcome. Short-term was defined as an outcome that took less than six months to occur. Long-term was defined as an outcome that took more than six months to occur. Unfortunately, many participants misunderstood that the short-term and long-term ratings were meant to be mutually exclusive. Only 10 of the total 21 participants provided complete data for the 89 ratings. The remaining participants (a) rated some of the 89 statements as *both* short-term and long-term or (b) left multiple statements blank or (c)

recorded ratings in an alternating pattern with every other statement marked as short-term (which suggested a misunderstanding of the instructions). Two participants explained that they were not familiar enough with the details of some outcomes to be able to accurately rate how long the outcomes took to occur. Thus, the proportion of long-term outcomes for each cluster is based on partial data.

Finally, Table 1 displays information which characterizes the entire group of participants in terms of experience with community organizing, personal education level, age and childcare responsibilities. The survey used to collect this information (which

Table 1

Participant Characteristics for Brainstorming and Sorting/Rating Groups

Characteristic	Brainstorming Group (n=19)	Sorting/Rating Group (n=21)
Experience with MOP		
1 year or less	21%	5%
2-3 years	21%	33%
4 or more years	58%	62%
LOC affiliation		
Faith community	68%	72%
School community	42%	33%
Education level		
No HS diploma	16%	14%
HS diploma	37%	33%
2-year degree	5%	5%
4-year degree +	37%	43%
Missing	5%	5%
Age		
Under 35 years	16%	19%
35-50 years	47%	33%
Over 50 years	37%	48%
Caring for children?		
Yes	47%	38%
No	53%	62%

Note: One of the twenty participants who brainstormed did not complete a survey. Percentages of LOC affiliation do not add to 100% because some participants are affiliated with both a faith and school LOC.

was available in English or Spanish) is presented as Appendix F.

Representation of Statements – Concept Mapping Analysis

The first step in representing the data is to run a multi-dimensional scaling analysis which puts the statements on a map. The map is shown as Appendix G. The goodness-of-fit measure for an MDS solution is called the "f-stress" or simply the "stress" (Kruskal & Wish, 1978, p. 24, 27; Ripley, 2011). Kruskal and Wish provide a discussion of the stress measure and they offer the rough guideline that if stress is greater than 0.10, then the number of dimensions chosen for the MDS solution may not be correct (Kruskal & Wish, 1978, p. 56). In the case of a concept map, a high stress value may indicate that two-dimensions are not adequate for presenting the data and that more dimensions are needed. The stress value for the two-dimensional solution for these data is 0.24, which is certainly higher than ideal.

Kruskal and Wish also note that conditions for interpreting stress include a lack of ties in the data (1978, p. 53). However, avoiding ties in the data is not possible in the concept mapping process. With eighty-nine statements and only twenty-one sorters, there are many pairs which are sorted together the same number of times. The presence of ties in the data can be seen from the Shepard diagram (Shepard, 1962a, 1962b) in Appendix H. The Shepard diagram takes each pair of the 89 statements ('89 choose 2' or 3916 pairs) and plots the input dissimilarity on the x-axis and the MDS configuration distance on the y-axis. The vertical alignment of points makes it evident that the MDS analysis must assign differing distances to many statement pairs which have the same dissimilarity measure.

Next, the statements on the MDS map were clustered using the four methods which were described above. Each map was clustered into 4, 5, 6 and 7 groups. The groups of cluster maps are shown as Appendices I, J, K and L with ellipses and varied plot characters demarking the clusters. The AGNES algorithm (Appendix I) provides a measure called the agglomerative coefficient which quantifies whether or not there is a natural cluster structure in the data. The agglomerative coefficient ranges between 0 and 1 with higher values indicating a clear clustering structure (Kaufman & Rousseeuw, 1990, p. 213). For the AGNES solutions, the agglomerative coefficient was very high with a value of 0.98. The DIANA algorithm (Appendix J) produces a similar measure ranging between 0 and 1. Its value is 0.93. Kaufmann and Rousseeuw note, however, that both the agglomerative and divisive coefficients can be influenced by even one outlier. Based on a visual inspection of the maps, a tight clustering structure is not evident. Therefore, the high values of the coefficients may be more attributable to outliers rather than clear clustering structure. The FANNY (Appendix K) algorithm calculates a normalized version of Dunn's partition coefficient (J. C. Dunn, 1976) to assess the clarity of cluster structures produced (Kaufman & Rousseeuw, 1990, p. 171). This measure ranges from 0 to 1 with 1 indicating a completely well-partitioned (e.g. non-fuzzy) cluster solution. The normalized partition coefficients for the FANNY cluster solutions ranged from 0.133 to 0.214, indicating a set of rather poorly-partitioned cluster solutions. Poor partitioning measures are validated by a visual inspection of the map. With the exception of a well-defined cluster at the left of the map, the rest of the clusters are not well-differentiated. The PAM algorithm (Appendix L) does not offer a numeric measure of the goodness of the clustering solution.

After running the sixteen clustered solutions, the next step was to narrow the set to some reasonable number of maps which could be interpreted by the participants. One of the methods appeared to break down at the 7-group level. FANNY did not produce a 7-group solution; its "7-group" solution contained six clusters which were virtually identical to the 6-cluster solution. Also, two of the 7-group solutions (DIANA and AGNES) had one cluster which was very small, with either two or four statements, respectively. In the context of theory-building, it seems that small, specific groups might not be helpful. Therefore, in the interests of keeping interpretation meaningful, and, since not all of the 7-cluster solutions were viable, the 7-group solutions were eliminated from consideration. Next, the remaining DIANA solutions were eliminated because (a) the 4-group solution appeared to poorly differentiate a visually obvious cluster on the left side of the map and (b) the 5- and 6-group DIANA solutions also displayed one cluster with only two members. These exclusions narrowed the field from sixteen possible cluster maps to nine.

The remaining nine cluster solutions displayed a great deal of similarity. Each of them presented cluster solutions arranged in a more-or-less oval-shaped pattern around a relatively empty area slightly to the upper left of the center of the map. To compare the nine remaining maps, two cluster validation indices were calculated: the Dunn index and the Davies-Bouldin index (Davies & Bouldin, 1979; J. C. Dunn, 1974; Halkidi, Batistakis, & Vazirgiannis, n.d.). Both assess the separation of clusters. These indices can be calculated in the R statistical environment using a number of different intracluster diameter and intercluster distance measures. Because the MDS map appears to contain both outlier points (e.g. statement 89) and overlapping points (e.g. statements 31 and 18),

average intracluster diameter and intercluster distance measures were used to calculate the indices. Average intracluster diameter is the average of all distances between the point pairs in a cluster (Nieweglowski, 2009). Average intercluster distance is the average distance between all possible point pairs formed by taking one member of the point pair from one cluster and the other member of the point pair from another cluster (Johnson & Wichern, 2007, p. 681; Nieweglowski, 2009). Other options are complete linkage (for both intracluster diameter and intercluster distance) and single linkage (for intercluster distance). These later measurements, however, are based on the furthest apart and closest together point pairs, respectively. Such calculations may be unduly influenced by outlier or overlapping points. Average linkage distances are used so as to be more representative of the entire cluster.

Dunn's index is a measure of dissimilarity between clusters. Thus, one is interested in finding the cluster solution with the highest value for Dunn's index. For the nine remaining solutions, the values of Dunn's index ranged from 1.54 to 3.04. The solution with the highest index (3.04) was the PAM 4-cluster solution. However, the index for the AGNES 5-cluster solution was virtually identical at 3.00. The Davies-Bouldin index is a measure of similarity between clusters. Thus, one is interested in finding the cluster solution with the lowest value for the Davies-Bouldin index. For the nine remaining solutions, the values of Davies-Bouldin ranged from 0.47 to 0.56. The solution with the lowest index was the AGNES 4-cluster solution.

Two maps were presented at the interpretation session with MOP leaders: these are the AGNES 4- and 5-cluster solutions. In addition to the statistical support noted above, there are two other reasons that these solutions are good choices for interpretation.

First, three of the groups are identical in the AGNES 4- and 5-cluster solutions. Furthermore, the 5-cluster solution breaks the least cohesive cluster from the 4-cluster solution in two. It seems that this area of the map warrants further scrutiny. Presenting both solutions would allow participants to name the groups individually or as one. Second, the AGNES algorithm appears to do a better job than PAM of differentiating the visually obvious cluster at the left side of the map. Thus the AGNES solutions were chosen for interpretation.

Interpretation of Maps

The interpretation session was held on April 6th, 2011, from noon to 1:30 PM at the MOP offices. Only three previous participants from MOP attended. MOP's executive director also attended the meeting and participated in the discussions. The session began by reviewing the concept mapping process for the group. Then the AGNES 4-cluster solution (shown in Appendix M) was displayed for everyone to see and participants reviewed a list of which statements were included in each of the four clusters. The group sought to come up with a name for each of the four clusters.

The interpretation participants were able to characterize clusters 1 through 3 quite quickly. Cluster 1 (at the left of the map in Appendix M, with circular plot characters) was named *victories*. This name signifies the public outcomes resulting from MOP's organizing work. Cluster 2 (at the top right of the map with triangular plot characters) was named *personal development*. This name signifies (a) newly-acquired skills that benefit MOP leaders in their personal lives and (b) feelings of personal empowerment that leaders gain when they participate in the organizing process. Cluster 3 (with cross-

shaped plot characters) was named *public leadership skills*. This name signifies the political and democratic process skills that leaders learn and use when they work to address community problems.

The participant group noted that cluster 4 (with x-shaped plot characters) was "the trickiest one" and also "tough". This was not surprising, as cluster 4 is visually the least tight and well-defined of the clusters. The group suggested that many of the statements deal with building "power" to effect community change. As this discussion continued, the AGNES 5-cluster solution was also displayed and participants reviewed a list of the statements from cluster 4 – now split into two groups (see Appendix N). The revised cluster 4 (still with x-shaped plot characters) is a small group at the lower left of the map. Cluster 5 (diamond-shaped plot characters) is in the central lower half of the map. The smaller, revised cluster 4 was now named *relationships with power people*. The participant group still struggled to name cluster 5. (Someone wondered skeptically how the computer had put this group together!) In their discussion, the participants mentioned several characteristics of the statements in cluster 5: (a) it describes "how you work within MOP", (b) this group's statements could form a "how-to manual" of MOP's activities, (c) these are the things a person needs in his/her "heart" if s/he is a leader at MOP and (d) many of the cluster 5 statements are about "involvement". One participant felt that two statements (50 and 57, both about relationships) definitely belonged with cluster 4 and not with cluster 5. Since the group was having trouble, the study co-PI (who was facilitating the discussion) suggested "MOP's culture" as a possible cluster name because cluster 5 seems to contain many statements about the behavior of MOP leaders; it is a list of what they do. The "culture" label seemed to resonate with some

group members; one person said that it was “very helpful”. Even at the end of the discussion about clusters 4 and 5, however, there was not unanimous agreement about whether to characterize these points as one or two groups and what to name cluster 5.

When creating the final map in Appendix N, the 5-cluster solution was used because it reduces the number of points that are difficult to characterize (i.e. new cluster 5 is smaller than old cluster 4). During subsequent analysis, cluster 5 was renamed *culture of civic engagement* in an effort to reflect the fact that this cluster discusses many of MOP’s leaders’ activities. All of the other clusters are named as discussed above. Average timeframe ratings are also on the final map. Recall that each participant was asked to label each of the 89 statements as “short-term” or “long-term”. Short term statements are coded with a value of 0 and long-term statements are coded with a value of 1. Averaging across participants for each statement measures the proportion of respondents who rated that statement “long-term”. Averaging these proportions across the statements in each cluster produces a measure of the proportion of statements in the cluster which are characterized as “long-term”. The resulting timeframe measure for each cluster ranges between 0 and 1 with larger measures reflecting a larger proportion of statements designated as “long-term”. The final map with both cluster labels and timeframe ratings is shown in Appendix N. A list of grouped statements is shown in Appendix O.

Chapter 5 – Discussion

In this final chapter, the study's research questions are answered, suggestions for further study are made and additional questions raised by the study are discussed.

Research Question 1: Dubin's Units of Theory

The first research question asks: "How well does concept mapping assist in developing the outcomes domain of a program theory for a complex social intervention?" To answer the "how well?" research question, several of Dubin's (1978) distinctions are explored. The first is: "Does concept mapping produce potential units for incorporation into theory which (a) describe properties of objects instead of objects themselves and which (b) do not describe one-time events?" The second is: "Does concept mapping produce potential units which meet Dubin's four sets of mutually exclusive distinctions: attribute/variable, real/nominal, primitive/sophisticated and collective/member?" Each of the distinctions is addressed in turn, before drawing a conclusion to the overall research question.

Properties of objects instead of objects themselves. Dubin provides an example of distinguishing objects or "things" (p. 41) from properties of such objects or things. He notes that people or cities might at first glance appear to be what a researcher would study (Dubin, 1978, p. 40). But then Dubin suggests that things can be quite complex when taken as a whole; thus researchers might be better equipped to study a few properties or characteristics of many things, rather than the things themselves. Furthermore, if one is interested building potentially generalizable theories, one may not

be interested in specific people or cities. Rather, one would be interested in properties or characteristics which apply to many people or cities.

In the current study, the list of 89 organizing outcomes is a list of objects or things. Using concept mapping has enabled an enumeration of properties of the 89 things. The cluster names on the final map (Appendix N) do present themselves as properties of these things. Some outcomes on the list have the property of being *victories*. Some of the outcomes have the property of being instances of *personal development* or of *public leadership skills*. And some have the property of being *power relationships* or examples of *civic engagement*. Concept mapping is in this instance a useful tool allowing movement from the study of things (i.e. individual outcome statements), which are not easily incorporated into theory, to the study of properties, which are more useful for building theory.

Units of theory cannot be one-time events. The grouping of statements on the concept map also allows movement away from the description of one-time events. Dubin notes that units of theory cannot be one-time events. Some of the statements generated by the brainstorming process do reflect one-time events. Examples are "MOP helped get Salud health clinic in Commerce City. It now has its own building and soon 10 more doctors" and "MOP/PICO helped us win the hospital provider fee bill (100,000 new insured)" (see Appendix C, items 18 and 22). Some of the statements generated by the brainstorming process reflect more general occurrences which are not one-time events. Examples are "I get to know people I would otherwise not have known" and "MOP puts a 'face' on issues and data, humanizes problems in the community" (see Appendix C, items 5 and 21). Finally, some statements reflect a combination of both one-time and

more general happenings: "MOP established Parent Liaisons in schools; PLs help connect us to resources in the school" (Appendix C, item 14). The establishment of parent liaisons is a one-time event, at least within a given school. But the phenomenon of parent liaisons connecting "us" (i.e. other parents) to resources in the school is presumably a more general happening.

When one examines not the individual statements, however, but rather the clusters on the map, one finds that none of the clusters represent events resulting from a specific set of historical circumstances which occurred only once. *Victories* occur repeatedly over time. *Personal development* and *public leadership skills* are on-going. *Power relationships* and *civic engagement* also do not occur in just one instance, they occur in different contexts and at different times. The nature of each cluster is that its occurrence can be examined (a) in the context of different issue campaigns at MOP, (b) at different times and, potentially, (c) even in the context of another similar community organization which is not MOP. Thus, in addition to allowing movement from the study of things to the study of *properties* of such things, concept mapping also moves the researcher or theory builder from consideration of one-time events to consideration of more general events which do not occur only once. This is in keeping with Dubin's criteria for good theory.

Attribute or variable. The fact that cluster names on the map *are* properties and that they *are not* one-time events suggests that they are promising candidates to serve as units for theory-building. However, Dubin (1978) also presents four other dichotomies relevant to theoretical units. He notes that theory units must have one characteristic from each of the four pairs. First, Dubin argues that units of theory must be either attribute or

variable. Dubin stipulates this requirement in order that a researcher may make observations or measurements as part of the process of confirming or disconfirming a theory. Recall that an attribute is a property which is either present or not present in the object which is under study. A variable, in contrast, is a property of some object that may be present in degree (Dubin, 1978). In other words, an attribute is a characteristic which is either present or not in any object, while a variable is a characteristic which is present at varying levels.

Examining the clusters from the final concept map shows that some of them have the potential to be characterized as attributes. *Victories* and "power relationships" are examples of such. A public meeting or an issue campaign might be characterized as a victory or not. Similarly, for any public official in Denver, one could specify whether that person has a relationship with MOP or not. Leaders' personal development and their public leadership skills can be considered variable units of theory. At a later stage of theory-building, a researcher might discover or develop empirical indicators for these properties so that they could measure whether individual leaders at MOP or at other community organizations have a high or low level of personal empowerment or public skills. Of all the clusters on the map, *culture of civic engagement* is the most difficult to characterize as either attribute or variable. However, it might be possible to measure MOP's level of success in building an organizational culture of civic engagement, or to measure MOP members' civic engagement in relation to that of other people who are not MOP members. This examination of the clusters shows that concept mapping produces units which meet another of Dubin's (1978) theoretic criteria.

Real or nominal. Dubin's distinction between real and nominal units "…rests only on the ability of the scientist ultimately to secure empirical indicators of the units he employs in his theories…" (Dubin, 1978, p. 51). At this point in the theory development process for MOP, the units represented by clusters on the map are nominal; no exact empirical indicators are available. However, with further work, one might expect that empirical indicators could be defined, thus changing the units from nominal to real and making them useful for the empirical testing of theory. For example, a definition of "victory" might be developed so that each issue campaign could be classified as a victory or not. Or, the number of people at a public meeting might be deemed an indicator of civic engagement and be useable to measure MOP's success with building a culture of civic engagement. Finally, leadership skills and personal empowerment are both previously studied concepts. Measurement scales from the fields of human resource development and psychology might be available as empirical indicators for these units. The ability to classify potential units as nominal, and the evident possibility for converting these nominal units to real ones, displays again the utility of concept mapping for identifying units of theory.

Primitive or sophisticated. Recall from Chapter 2 that Dubin's third distinction is between a primitive and a sophisticated unit. This distinction is relatively simple to understand, as primitive units are basically undefined and sophisticated units are defined. Dubin uses the example of a factor analysis to illustrate the difference between a primitive and a sophisticated unit. A factor analysis uses observed values of variables and correlations between variables to deduce the presence of unobservable quantities or

factors (Johnson & Wichern, 2007). The unobservable and as-yet-undefined factors in such an analysis are examples of primitive units of theory.

The process of concept mapping involves both primitive and sophisticated units. Before naming, units (or potential units) are undefined. They exist only as numbered clusters on a map. However, the process of interpreting the concept map begins to define the units and moves them from primitive to sophisticated status. It is certainly plausible that the definitions for units which emerge from concept mapping are not final and will be further refined. However, interpretation during the concept mapping process provides theoretical units which are more sophisticated than primitive and thus meets another of Dubin's (1978) criteria for good units of theory.

Collective or member. Dubin's (1978) last distinction deals with the question of whether individuals (i.e. members) or classes (i.e. collectives) of individuals can be considered the basis for a theoretical unit. In this section, Dubin seems to lose sight of his first basic requirement that theoretical units be composed of properties of things and not the things themselves. He should more appropriately ask whether a unit of theory is a *property* of an individual or a *property* of a class. To be consistent with the prior definitions, one can consider whether the cluster names on MOP's map represent *properties* of individuals or classes.

The answer to this question is that most of the units appear to be more easily characterized as properties of individuals, rather than a collective or class. The *victories* unit seems to be a property of individual events, rather than groups of such events. *Power relationships* is applicable to individual public officials and both the *personal development* and *public leadership skills* units of theory describe characteristics of

individual leaders. Finally, the *culture of civic engagement* unit is more easily characterized as collective because it applies to the entire group of MOP leaders. In order to make multiple observations or measurements about this unit, one would need to extend the program theory beyond MOP and observe the culture at other community organizations. Characterization of potential units as collective or member shows that concept mapping is able to fulfill the fourth of Dubin's (1978) four either/or distinctions for good theory.

Concept mapping and program theory. The study's initial research question is: "How well does concept mapping assist in developing the outcomes domain of a program theory for a complex social intervention?" Using Dubin's framework as a guide, it is fair to respond to this question by saying that concept mapping appears to do the job well. In the case of Metro Organizations for People, the concept mapping process was able to take a list of 89 unorganized outcomes from a brainstorming session and organize them into five outcome groups which can represent units in a theory. The groups meet Dubin's criterion of being properties rather than things, and they also meet the criterion that theoretical units not be one-time events. Furthermore, the units represented by the clusters on the concept map can be characterized as sophisticated rather than primitive, but still nominal (instead of real) because they lack empirical indicators. Four of the units (excepting *culture of civic engagement*) appear to be fairly easy to characterize as attribute or variable and all of these four units are member units rather than collective ones. Thus, with the possible exception of the *culture of civic engagement* unit, concept mapping has provided good material for theory-building.

Research Question 2: Parsimony and Comprehensiveness

Whetten (1989) provides two useful criteria against which to evaluate the elements comprising a theory. These are parsimony and comprehensiveness. Essentially, parsimony asks whether there are any redundant, non-informative elements within a theory and comprehensiveness asks whether there are any pieces missing from the theory. In order to answer the question about comprehensiveness, participants at the interpretation session were asked whether they thought that any clusters on the map were missing. The group definitely believed that nothing was missing; they quickly arrived at this consensus. They were also asked if the groups were "really different" from each other, in order to answer the parsimony question. One person responded that some statements could be "moved around" to other groups. Then multiple participants observed that the steps of MOP's organizing model (Snyder, n.d.) were all represented on the map. Although the group did not respond as quickly and clearly to this question as they did to the "is anything missing?" question, they provided no evidence that they thought any of the groups were redundant. Thus, the second research question is answered affirmatively: concept mapping produces a set of theoretical units which are parsimonious and comprehensive.

Research Question 3: Validity

As noted in Chapter 3, there are only a few conversations in the literature about the validity or trustworthiness of concept maps. Since validity is a property of inferences or knowledge and not a property of methods, the last research question is phrased as follows: What evidence exists for the validity of programmatic outcomes as articulated by the concept mapping process? The current study explores two sources of validity

evidence. The first is statistical evidence generated by the multiple methods of clustering used. The second relies on comments made by participants during the interpretation sessions.

Evidence from multiple clustering methods. As described in Chapter 4, there were originally sixteen different cluster maps created. The DIANA cluster solutions and the 7-group cluster solutions were eliminated from consideration, which left a group of nine potential solutions from which to choose. The AGNES 4-cluster and AGNES 5-cluster solutions were presented to participants for the interpretation session.

All of the nine viable solutions appeared visually quite similar. This result lends some support to the notion of natural clusters in the data (Johnson & Wichern, 2007). Johnson and Wichern (2007, p. 28) note that Chernoff faces can be used to verify the goodness of a particular cluster solution. However, Chernoff faces can only be used when there is a multivariate observation for each object being clustered. In this case, there is only a brainstormed statement, not a multivariate observation. Instead of Chernoff faces, similarity indices were run (a) for all pairs of solutions sharing the same number of groups and (b) for all pairs of solutions using the same clustering method. The similarity index is based on a measure called partition-distance (Almudevar & Field, 1999; Gusfield, 2002). A partition for a set of objects is simply the division of that set into smaller, non-overlapping groups (Ross, 1988). For any two different partitions, the partition distance is the "minimum number of elements that must be deleted" from the set of objects so that the two partitions, when restricted to the remaining elements of the set, are equal (Giurcdneunu, Tabus, Shmulevich, & Zhang, 2003, p. 57). To calculate a similarity index, the partition distance is converted to a measure called an "assignment"

where the value of the assignment equals the number of objects in the set minus the partition distance (Giurcdneunu, et al., 2003, p. 57). The similarity index between two partitions is based on the value of the assignment and it ranges in value from 0 to 1, where a value of 1 indicates identical partitions (Nieweglowski, 2009).

The context for use of the similarity index in Giurcaneanu, et al., is a comparison of multiple cluster solutions (with different numbers of clusters) to a "true" partition for a microarray dataset (2003, p. 57). The current study departs from the context in which the similarity index was developed. As shown in Table 2, the similarity index is here used to

Table 2

Similarity Indices Comparing Different Cluster Solutions

Comparison Pair	Similarity Index
Comparisons with 4 clusters	
AGNES 4 vs. PAM 4	0.90
AGNES 4 vs. FANNY 4	0.92
PAM 4 vs. FANNY 4	0.90
Comparisons with 5 clusters	
AGNES 5 vs. PAM 5	0.91
AGNES 5 vs. FANNY 5	0.88
PAM 5 vs. FANNY 5	0.91
Comparisons with 6 clusters	
AGNES 6 vs. PAM 6	0.90
AGNES 6 vs. FANNY 6	0.75
PAM 6 vs. FANNY 6	0.74
AGNES Comparisons	
4 vs. 5 Clusters	0.90
4 vs. 6 Clusters	0.84
5 vs. 6 Clusters	0.94
FANNY Comparisons	
4 vs. 5 Clusters	0.81
4.vs. 6 Clusters	0.73
5 vs. 6 Clusters	0.80
PAM Comparisons	
4 vs. 5 Clusters	0.78
4.vs. 6 Clusters	0.73
5 vs. 6 Clusters	0.89

compare eighteen pairs chosen from the nine cluster maps and to assess how similar are the pairs. As can be seen, all of the 18 comparisons show a great deal of similarity between cluster solutions; none of the indices are lower than 0.73.

The use of multiple clustering solutions yields nine cluster maps which are visually similar and which are quantitatively similar as measured by the similarity index. These results lend clear support for the validity of the final cluster map. One can be confident that the theoretical units resulting from the map are not an artifact of the clustering solution used. It is reasonable to expect that if different cluster maps had been presented at the interpretation session, similar group names would have resulted.

Interpretation session results. The second source of evidence for the validity of programmatic outcomes as articulated by concept mapping relies on comments made by participants during the interpretation session. In the following paragraphs, participant comments are analyzed. This analysis suggests that the concept map is a valid representation of MOP's outcomes.

The first cluster is named *victories*, and it represents the public outcomes of MOP's work. One participant referred to this group of outcomes as a "pillar" of MOP's work. He noted that if the only outcomes from MOP's work were those in the other clusters (i.e. *personal development, public leadership skills, power relationships* and *culture of civic engagement*) then MOP should be "out of business" because it would never effect any changes in the larger community. Conversely, if MOP only achieved victories and did not develop people, then it would be an "advocacy" organization, not a community organizing group. The participant concluded that some "fundamental aspects of who we are" are "beautifully" listed in the mapping results.

Another participant commented on the relationship between cluster 2 (*personal development*) and cluster 3 (*public leadership skills*). Note that these clusters are very close together on the map; they lie mostly in the top, right quadrant. The observation was made that it is reasonable to have these clusters close together on the map because personal development is required before leadership development can happen. The same participant also suggested that the personal development outcomes might be stated primarily by newer leaders at MOP while the leadership development outcomes would be stated by more experienced leaders. Another participant followed up by saying that he was not surprised to see the *public leadership skills* cluster, because teaching "public skills" is one of the main goals for MOP's staff of organizers. However, the *personal development* cluster is also very important; in fact, this topic had been an item of discussion at a meeting with one of MOP's funders on the day prior to the interpretation session. Another participant summed up the importance of the *personal development* cluster by saying that being with MOP "makes me stronger". Finally, a closing comment about the first three clusters was that "those three are really clear".

Several final comments were also helpful for validating the map. One person observed that the statement list and the clustered outcomes were "all I would need" to explain to someone what MOP is about. And, furthermore, that it was "impressive" that the cluster groupings could represent characteristics of many different MOP leaders, not just one. Another participant said that she saw "all the steps" of MOP's organizing process in the map. Someone else described the map as "driven by an overall organizing model," even though no brainstorming participants actually named the model in any of the statements. The model reemerged from the concept mapping process. When

participants were asked *how* the model was reflected in the map, the response was: "It's *everything*". One speaker went on to mention the following list of aspects of MOP's organizing model which are reflected in the map: the development of people, the development of politics and power skills, taking action, leaders leading, building relationships with people in power, ordinary people getting involved in politics and "winning".

One of the timeframe measures also lends some additional evidence of validity to the final map. Note that the timeframe measure for the *victories* cluster is 0.9. This indicates that the participants who rated the statements in this cluster are in agreement that victories are long-term outcomes which take more than six months to achieve. The remaining clusters all have timeframe measures of either 0.5 or 0.6. This indicates that participants are not in agreement about whether the statements comprising the other clusters were short-term or long-term outcomes. Apparently this assessment varies with people's individual experience. However, what is noteworthy is that the clusters with shorter timeframe ratings are all related to outcomes which necessarily *precede* achieving public victories. MOP seeks to develop people, to build a culture of civic engagement within itself and to develop relationships with power players before it can achieve public victories. These logical timeframe ratings provide support for the notion that the map validly reflects MOP's work.

As an additional validation, the results of the concept map were compared to information provided on MOP's website in the section entitled *Accomplishments and Successes* (Metro Organizations for People, n.d.b). This was done to determine whether the information provided on the map is mirrored in MOP's own public description of

itself. As MOP discusses its work on the website, activities which fall under four of the five clusters are mentioned. Public accomplishments (i.e. *victories*) which are mentioned include: a neighborhood transit plan in Sun Valley, a discipline model and parenting classes at Cole Academy of Arts and Sciences, traffic safety signs at Harrington Elementary and a student-driven teacher evaluation system at West High School. *Public leadership skills* receive many mentions on MOP's website, including: conducting at least seven recent public meetings, and training over 400 community leaders. The *relationships with power people* cluster is reflected by the fact that several public officials attended MOP's various meetings: Colorado Senator Michael Bennett, the Colorado State Legislature Speaker of the House, representatives from the Governor's office and professors and administrators from the University of Denver. The *civic engagement* cluster is also mirrored on the website with mention of press conferences and voter registration drives. The only cluster which is not clearly mentioned under *Accomplishments and Successes* is the *personal development* cluster. However, a banner at the bottom of MOP's home page provides a profile of five active MOP leaders (Metro Organizations for People, n.d.c). The leader profiles on the website mention several instances of personal development. These include a life-changing opportunity to testify before a congressional committee (for one leader) and another youth leader's experience of having peers who now look to her for advice. Thus, the information on the concept map produced by MOP leaders appears consistent with information and self-description provided on MOP's website.

Taken together, participant comments, timeframe ratings and the material on MOP's website lend a great deal of supporting evidence for the validity of the concept

mapping output. To understand how this is so, it may be helpful to borrow some terms from naturalistic inquiry. Naturalistic inquiry often occurs in a constructivist paradigm (Denzin & Lincoln, 2005), which certainly contrasts with the critical realistic paradigm guiding this study. However, as the comments from MOP leaders are clearly qualitative data and as much of naturalistic inquiry is also qualitative, such borrowing may be appropriate as we seek to establish the validity of a map based in part on qualitative comments. Lincoln and Guba (1985) note in their text, *Naturalistic Inquiry*, that four criteria may be used to establish the "trustworthiness" (p. 294) of research results. These criteria are: credibility, transferability, dependability and confirmability. In a naturalistic study, these four trustworthiness criteria replace four traditional research criteria: internal validity, external validity, reliability and objectivity. (Lincoln & Guba, 1985, pp. 294-301). Since the third research question concerns the internal validity of the maps (i.e. how well do they represent MOP's programmatic outcomes), the relevant concept to borrow from naturalistic inquiry is that of *credibility*. To establish credibility, "...the naturalist must show that...the *reconstructions*....that have been arrived at via the inquiry are *credible to the constructors of the original...realities*" (Lincoln & Guba, 1985, p. 296). In the context of this study, it must be shown that the outcomes reconstructed on the concept map are credible to MOP leaders. Based on the comments discussed in the previous paragraphs, it is clear that both the leaders and the executive director who were part of the interpretation session recognized the reality of their organizing work in the output of the map. Furthermore, the interpretation session itself is an example of a "member check" (1985, p. 314), which is one of the five methods Lincoln and Guba suggest for establishing the credibility of a naturalistic research study. Leader comments

and the interpretation of these comments through a naturalistic framework suggest that concept mapping is a good tool for producing a credible and valid articulation of a collective concept.

Delimitations and Limitations

As noted in the introduction, this study is delimited in two considerable ways. First of all, it provides positive evidence for the usefulness of concept mapping for theory-building in only one program theory domain. The fact that results are limited to the outcomes domain also implies that they do not encompass the causal-explanatory portion of a theory. Inherent in this delimitation, however, is the opportunity to extend the work begun here and explore whether or not concept mapping is also useful in some or all of the remaining five domains which are part of Chen's (1990) framework (see Figure 1). Particularly of interest, would be whether concept mapping can be useful in articulating causal mechanisms that link treatments and outcomes. Secondly, the study is also delimited in that it examines only one group of participants at one community organization in one city. To build further evidence in support of using concept mapping for theory-building, additional studies in other contexts should be done (Pawson & Tilley, 1997). To address this second limitation, the question to answer (via further research) is whether theory articulated for Metro Organizations for People would be transferable (Lincoln & Guba, 1985, p. 297) to other community organizations.

Some additional limitations emerged as the study progressed. First, the number of participants was lower than what would have been ideal. MOP staff tried to recruit about thirty participants, but only twenty participants attended the brainstorming sessions. Because not all of the twenty participants returned sorting packets, additional sorters

were recruited at a MOP training weekend. Ideally, the study would have had more people sorting the statements. More sort solutions might have been helpful to better characterize the statements currently in the *civic engagement* cluster (which was difficult for participants to name). This cluster is near the center of the map and the statements comprising it do not form a tight cluster. Such placement suggests that the statements in the *civic engagement* cluster were not consistently sorted together with any other group of statements. They lie at the center of the map because they were not easily characterized and there was likely much variation in the way they were sorted. If more people had sorted all of the statements, the *civic engagement* cluster statements might have been better characterized. Twenty sorters are sufficient for statements that are relatively easier to characterize, as in the other clusters. But, for statements which are more difficult to characterize, more sorters are needed to allow a more meaningful clustering to emerge.

Another limitation to the study might be characterized as "participant fatigue" in relation to the concept mapping process. Three participants commented that the sorting took a great deal of time. Sort time could be reduced by limiting the number of statements. As a process improvement, it would be worth considering how a large list of statements (there were originally 125) could be further reduced so that it would still be representative of programmatic outcomes but would also be easier to sort. Eighty-nine statements appear to be too many. Kane and Trochim (2007) suggest borrowing methods from content analysis to assist in limiting the number of statements. Using content analysis to further reduce the set might have reduced participant fatigue and generated tighter clusters. Also, the small number of participants (three leaders and one staff

member) at the interpretation session is evidence that participants were a bit burned-out by the process. The staff member at the interpretation session said many other participants had been invited, but most declined to attend. It was unclear whether leaders declined because they had recently attended several other MOP meetings or whether they simply felt that they had devoted enough time to the research project. Although it might have been unwieldy to have all prior participants at the interpretation session, it would have been better to base conclusions from the session on the opinions of more than four people. The evidence of participant fatigue with the process suggests that users of concept mapping should look for ways to ensure that it is not unnecessarily burdensome to participants.

Further Research

All studies highlight opportunities for further research and this one is no exception. Two important areas for further research are noted below.

Theory-building research. Recall from Chapter 1 that Chen's (1990) framework for program theory posits six domains: inputs, intervening mechanisms, impacts, outputs, implementation environment and generalization (see Figure 1). This study explores the utility of concept mapping for theory-building in only one of these domains: outcomes. Further research should explore whether or not concept mapping can articulate units of theory (Dubin, 1978) for other domains. For example, concept mapping may be successful in articulating inputs. Such an exercise could start with a prompt similar to that used in the current study, for example: "Think about your entire experience with MOP. When I participate in MOP's community organizing, this is what I do: _____." Developing a concept map of inputs for MOP's program theory

should include staff as well as leaders, since staff is the source of training and coaching for leaders. Establishing which domains are best suited to articulation via concept mapping is certainly an open topic for research.

Another methodological question related to both concept mapping and theory-building is whether additional steps could be added to the concept mapping process to assist in articulating program theory more clearly. In this study, the *civic engagement* cluster was the most difficult for participants to name, and a few of the statements in that cluster appear to perhaps belong better in other clusters on the map. Further research could investigate whether a second round of sorting would help clarify the *civic engagement* cluster. For example, participants could be asked to re-sort only the statements in the *civic engagement* cluster and then to name any new sub-clusters which result. This might yield new information or suggest that new sub-clusters belong with the other four clusters on the map. However, a re-sorting exercise would be logistically difficult. The MDS and cluster analyses would have to be repeated for the smaller subset of data. Then the map of sub-clusters emerging from the original *civic engagement* cluster would have to be merged with the larger map. However, if these logistical issues could be resolved, a re-sorting exercise might help clarify the program theory being articulated by producing clusters which are more easily named and which clearly meet the criteria for theoretical units from Dubin's framework (1978).

Applied statistical research. A second area that would benefit from methodological research is the question of whether it is reasonable to always limit the results of multi-dimensional scaling to two dimensions. Trochim (1989) noted that two dimensions are used to provide a map that participants can visualize. However, although

convenient, it is certainly not the case that the two dimensional MDS solution will always be the best one from a statistical standpoint. Good options could be developed for presenting results to participants for a higher-dimensional solution. For example, clustered groups of statements could be presented without a map. Also, computing capacity has increased enormously since 1989 when Trochim published the first concept mapping article. Additional graphical displays could be explored which would allow a three-dimensional solution to be presented to participants in a meaningful way.

There is some evidence that this study could have benefited from a higher dimensional map. As noted above, the stress value for the two-dimensional solution was 0.24. The stress value for a three-dimensional solution was 0.17 – a 29% reduction. After running a three-dimensional MDS solution in R, one can calculate distances between the points in three-dimensional space and then run clustering algorithms using these three-dimensional distances. What is not clear is how one would display the three-dimensional cluster solutions in a meaningful way. A comparison of the AGNES 5-cluster solution based on a two-dimensional MDS layout and the AGNES 5-cluster solution based on a three-dimensional layout yields a similarity index (Nieweglowski, 2009) of 0.69. The fact that this value is less than all of the similarity indices shown in Table 2 (which compare various two-dimensional solutions to each other) suggests that clusters based on the three-dimensional distances might have yielded somewhat different results. A more detailed analysis would be required to determine whether the names of five clusters based on a three-dimensional layout would differ appreciably from those based on the two-dimensional layout used in this study.

In addition to displaying the appropriate number of dimensions for an MDS solution, another research question that arises is whether MDS is suitable to portray a large number of paired objects using a small number of dissimilarity metrics. In the current study there are eighty-nine statements (i.e. sort objects). Pairing objects results in over 3900 possible pairs. For an MDS analysis, each pair requires a dissimilarity measure. Ideally, each pair would have a distinct dissimilarity measure that was different from all other pairs. However, with only 21 participants and a univariate dissimilarity measure (i.e. whether a pair was sorted in the same group or not), the dissimilarity measures for each pair are not distinct. In fact, the group dissimilarity matrix (GDM) contains (a) many zeros and (b) many ties. Such a situation is avoided in other contexts where there are (a) fewer pairs under consideration or (b) a more-nuanced, multivariate dissimilarity measure.

Texts which provide detail on the statistics behind multidimensional scaling do not offer any obvious guidance on the question of using MDS in a situation with many pairs and a relatively undifferentiated dissimilarity measure (Cox & Cox, 2001; Everitt & Rabe-Hesketh, 1997). Note, however, that two non-metric scaling examples given in Everitt and Rabe-Hesketh (1997, pp. 34, 43) use many fewer object pairs (60 and 84, respectively), and they present many fewer ties in the dissimilarity measures. Further research on the question of "many pairs and few dissimilarity measures" should start with a literature review highlighting studies which successfully use MDS. Such a meta-study might assist in setting guidelines for when MDS is useful in sparse data situations. Such a literature review is a better first step than a statistical simulation study. This is because the success of an MDS analysis ultimately hinges on the meaning inferred from the MDS

solution. It would be difficult to assess the quality of MDS solutions which were randomly generated as part of a simulation study. A literature review would provide the opportunity to review many studies in context.

Additional Questions Raised by the Study

In addition to the specific suggestions for further research which are discussed in the previous section, the study raises some additional interesting questions to explore. These are highlighted in the following paragraphs.

An alternative interpretation of concept mapping results. The most straightforward interpretation of the study's results is discussed above, namely, that the study sought to delimit itself to building theory for Chen's (1990) outcomes domain *and* that the results indeed stayed within that domain. However, another interpretation is possible. Even though participants were asked about outcomes only, it is possible that some of what they consider outcomes belong in program theory domains other than the outcomes domain. During the interpretation session, MOP's executive director commented that teaching public leadership skills is the main job of MOP's staff. This comment suggests that training activity may constitute a significant piece of MOP's treatment domain. Reconsidering the *victories* cluster, including its long-term timeframe rating, we can conclude that the victories do indeed belong in the outcomes domain. However, the other four clusters may belong in the intervening mechanism domain. It is plausible that (a) personal development, (b) acquisition of leadership skills, (c) an organizational culture of civic engagement and (d) development of relationships with people in power are all mechanisms by which the causal effect of leadership skill training is transmitted to victorious outcomes. This interpretation is supported by the fact that the

four potential intervening mechanism clusters all have timeframe ratings which are substantially shorter than the timeframe rating for the *victories* cluster (see Appendix N). Using Shadish, Cook and Campbell's language, these four clusters may be mediator variables (2002, p. 11). Clearly more validation work would need to be done to support this interpretation. However, if this conclusion is correct, it suggests that a researcher must be keenly aware that the concept mapping tool will not always yield exactly what was intended as the topic of study. In this study, concept mapping may have yielded more than just information for the outcomes domain.

Other frameworks for articulating program theory. The alternative interpretation in the paragraph above suggests taking another look at the frameworks used for theory-building. Clearly, Chen's (1990) multi-domain framework is useful. By considering additional domains of program theory, one can arrive at a plausible alternative interpretation for the study results. However, other frameworks for identifying necessary elements of program theory may also be useful and should be explored. Such domains come from the world of evaluation practice more than from the academic literature. For example, a simpler framework based on logic modeling uses a linear representation of "inputs", "outputs" and "outcomes" to lay out the theory behind a program (University of Wisconsin-Extension, 2002, p. 14). The W. K. Kellogg Foundation (W. K. Kellog Foundation, 2001) also provides similar resources for logic-modeling that rely on a framework with fewer domains than Chen puts forth. It may be the case that simpler frameworks would suffice for building theory specific to a small number of organizations. But Chen's additional domains of *implementation environment*

and *generalization* are likely needed if the transportability (Lincoln & Guba, 1985) of specific, local theory is to be investigated.

Whetten and an alternative interpretation. In addition to reconsidering Chen's (1990) domains, a closer look at Whetten's (1989) framework for more general theory-building is in order. As discussed in the introduction, Whetten's complete framework for a theory requires four elements: what, how, why and who/where/when. The initial interpretation of the study's results situates all of the theoretical units from the concept map in Chen's outcomes domain. This corresponds to identifying several of the "what" elements of a theory in Whetten's framework. Under the alternative interpretation of results, however, some of the clusters may belong in Chen's intervening mechanism domain. Using Whetten's criteria, units from the intervening mechanism domain belong in the *how* category of theory-building. Whetten states that the factors constituting the *what* of a theory must be related to each other and that the specification of relationships "...typically introduces causality" (1989, p. 491). Using Whetten's framework provides some further evidence that concept mapping may provide more information than what is initially sought by the researcher. This study began by looking for elements of theory that are part of the *what* in Whetten's framework. Concept mapping provided some evidence not only for *what* but also for *how* as well.

More questions about concept mapping. Several more questions may be of interest in regard to the concept mapping tool. First of all, one may legitimately ask: Does a researcher really need to do all the work that concept mapping requires? Would a one-time focus group be just as efficient and provide similar results for theory-building? A very interesting methodological study would be to address the same substantive focus

statement with two different groups of participants. One set of participants would use a focus group process (Creswell, 2008, p. 226), and the other would use the concept mapping process. Substantive results could be compared between the two groups, as well as participants' experience of the two processes in terms of enjoyableness, fatigue or time burden.

Second, a question connected to the philosophy of social science research is also in order. There has been discussion in the education literature over the last couple of decades about the paradigmatic assumptions and beliefs which are made by researchers and how those beliefs and assumptions should and do influence the outcomes of research (Guba, 1990; Guba & Lincoln, 2005; Phillips & Burbules, 2000). Several of the most commonly recognized research paradigms are recently summarized by Guba and Lincoln (Guba & Lincoln, 2005, p. 195); other paradigms have also been elsewhere identified, including the critical realist stance (Bechara & Van de Ven, 2007; Bhaskar, 1975) used in this study. It has been suggested that some research paradigms are "incommensurable" with others (Guba & Lincoln, 2005, p. 198). Guba and Lincoln also note that the debate around commensurability has included discussion of whether, when and how to combine *methods* for research that are commonly associated with different paradigms (2005, pp. 200-201).

The issue of combining methods (or not) begs the following question: can a method like concept mapping (which includes both qualitative and quantitative elements) be acceptable for use in multiple research paradigms? Consider Guba and Lincoln's comments on the types of methods which are used within each paradigm. For example, the critical theory paradigm (which differs from the *critical realist* paradigm) uses

methods which are dialogic and dialectical. The constructivist paradigm uses methods which are hermeneutical and the participatory paradigm uses methods of "collaborative action inquiry" (Guba & Lincoln, 2005, p. 195). Now consider the elements of the concept mapping tool. Concept mapping is dialogical in the sense that participants discuss amongst themselves which statements should be included and participants may work together to edit the statement list before it is sorted. Concept mapping is hermeneutical because the sorting exercise and the statistical analysis seek to draw larger meaning from the unorganized set of textual statements. And, finally, concept mapping is a prime example of collaborative research, where the experience and knowledge of participants complements and informs the expertise of the researcher. In addition, concept mapping may also be acceptable to researchers who subscribe to a postpositivistic research paradigm. Although the methods of postpositivism may be primarily quantitative, such research can include qualitative techniques as well (Guba & Lincoln, 2005). The preceding suggestions about the multi-paradigmatic usability of concept mapping are tentative. A further step in evaluating the ability of concept mapping to function in multiple paradigms would be to ask researchers who subscribe to the various paradigms to write about whether concept mapping could be used within their preferred paradigm and why.

Finally, related to the paradigmatic reflections above, one might also ask whether concept mapping belongs to the domain of *mixed methods* research. Teddlie and Tashakkori (2010) discuss nine general characteristics of mixed methods research. These are: (a) methodological eclecticism, i.e. the thoughtful combination of methods to best answer a question of interest, (b) paradigm pluralism, i.e. "...the belief that a variety of

paradigms may serve as the underlying philosophy for the use of mixed methods" (2010, p. 9), (c) an emphasis on diversity and on divergent results, (d) an emphasis on continua, rather than either/or dichotomies, (e) an iterative approach to research, (f) allowing the research questions to determine the methods used in a study, (g) possession of a set of *mixed methods* research designs and processes, (h) a valuing of balance and compromise in the research community and (i) use of visual representations to illustrate research design, data collection and analysis techniques. While concept mapping may not clearly meet all of these characteristics, it certainly does meet some of them. Concept mapping is definitely an example of methodological eclecticism (characteristic [a]), as it combines brainstorming techniques with multidimensional scaling and cluster analysis to achieve the specific purpose of articulating a collective concept. Concept mapping also appears to be a method that can live in different research paradigms (characteristic [b]), as the discussion in the previous paragraphs suggests. Concept mapping is already considered by some to be part of the set of mixed methods research processes (characteristic g; Ivankova & Kawamura, 2010; Kane & Trochim, 2007; Trochim & Kane, 2005). And, finally, it is amenable to use of a visual display (characteristic [i]) to depict the data collection and analysis methods (see Figure 2). While concept mapping may not directly provide an example of some of the other mixed methods characteristics, it is certainly not incompatible with other characteristics such as: an emphasis on diversity and/or continua, an iterative approach to research, a research question dictating method and a valuing of balance or compromise in the research community. Thus, it seems reasonable to conclude that concept mapping can fit comfortably within the evolving domain of mixed methods research.

Conclusion

In conclusion, concept mapping is a promising tool for articulating program theory. It offers a way of articulating outcomes from a "messy" social intervention. In the context of this study, it produced an articulation of outcomes that was a parsimonious, comprehensive and valid representation of the work of Metro Organizations for People in Denver. The outcomes articulated also meet the requirements laid out by Dubin (1978) to be useful as units in a systematic statement of theory. Concept mapping is a promising tool for theory-building. However, it is a tool which is perhaps not fully developed; it should be refined further and used more.

References

Almudevar, A., & Field, C. (1999). Estimation of Single-Generation Sibling Relationships Based on DNA Markers. *Journal of Agricultural, Biological, and Environmental Statistics, 4*(2), 136-165.

Argyris, C., & Schoen, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.

Argyris, C., & Schoen, D. A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison-Wesley.

Bechara, J., & Van de Ven, A. H. (2007). Philosophy of Science Underlying Engaged Scholarship. In A. H. Van De Ven (Ed.), *Engaged scholarship: A guide for Organizational and Social Research* (pp. 36-70). Oxford, England: Oxford University Press.

Bedi, R. P. (2006). Concept mapping the client's perspective on counseling alliance formation. *Journal of Counseling Psychology, 53*(1), 26-35. doi: 10.1037/0022-0167.53.1.26

Bedi, R. P., & Alexander, D. A. (2009). Using multivariate concept-mapping for assessing client understandings of counselling. *Canadian Journal of Counselling, 43*(2), 76-91.

Bhaskar, R. (1975). *A realist theory of science*. Leeds, England: Leeds Books.

Bhaskar, R. (1998). General Introduction. In M. Archer, R. Bhaskar, A. Collier, T. Lawson & A. Norrie (Eds.), *Critical realism: Essential readings* (pp. ix - xxiv). London: Routledge.

Bickman, L. (1979). Program evaluation and social psychology: The achievement of relevance. *Personality and Social Psychology Bulletin, 5*(4), 483-490.

Bickman, L. (1989). Barriers to the use of program theory. *Evaluation and Program Planning, 12*(4), 387-390. doi: 10.1016/0149-7189(89)90056-6

Bigne, J. E., Aldas-Manzano, J., Kuster, I., & Vila, N. (2002). The concept mapping approach in marketing: An application in the travel agencies. *Qualitative Market Research: An International Journal, 5*(2), 87-95. doi: 10.1108/13522750210423724

Birckmayer, J. D., & Weiss, C. H. (2000). Theory based evaluation in practice: What do we learn? *Evaluation Review, 24*(4), 407-431.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*(3), 296-322. doi: 10.1111/j.2044-8295.1910.tb00207.x

Brug, J., Steenhuis, I., Van Assema, P., & De Vries, H. (1996). The impact of a computer-tailored nutrition intervention. *Preventive Medicine*, *25*(3), 236-242. doi: 10.1006/pmed.1996.0052

Buja, A., Swayne, D. F., Littman, M. L., Dean, N., Hofmann, H., & Chen, L. (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics, 17*(2), 444-472. doi: 10.1198/106186008x318440

Cacy, J. R. (1996). *The reality of stakeholder groups: A study of the validity and reliability of concept maps.* (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (Publication No. AAT 9608493)

Carter, J. W., Enyedy, K. C., Goodyear, R. K., Arcinue, F., & Puri, N. N. (2009). Concept mapping of the events supervisees find helpful in group supervision. *Training and Education in Professional Psychology, 3*(1), 1-9. doi: 10.1037/a0013656

Carvalho, S., & White, H. (2004). Theory-based evaluation: The case of social funds. *American Journal of Evaluation, 25*(2), 141-160. doi: 10.1177/109821400402500202

Chen, H.-t. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.

Chen, H.-t., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces, 59*(1), 106-122.

Chen, H.-t., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*(3), 283-302.

Chen, H.-t., & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10*, 95-103. doi: 10.1016/0149-7189(87)90025-5

Cohen, B. P. (1991). *Developing sociological knowledge: Theory and method* (2nd ed.). Chicago: Nelson-Hall.

Concept Systems Incorporated. (2005). The concept system (Version 4.0). Ithaca, NY: Concept Systems Incorporated. Retrieved from www.conceptsystems.com

Cook, T. D. (2000). The false choice between theory-based evaluation and experimentation. *New Directions for Evaluation, 2000*(87), 27-34. doi: 10.1002/ev.1179

Cox, T. F., & Cox, M. A. A. (2001). *Multidimensional Scaling* (2nd ed.): Chapman and Hall.

Creswell, J. W. (2008). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (3rd ed.). Upper Saddle River, NJ: Pearson Education.

Davidson, E. J. (2000). Ascertaining causality in theory-based evaluation. *New Directions for Evaluation, 2000*(87), 17-26. doi: 10.1002/ev.1178

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1*(2), 224-227.

Denzin, N. K., & Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (3rd ed.). Thousand Oaks, CA: Sage.

Dubin, R. (1978). *Theory building* (Rev. ed.). New York: Free Press.

Dumont, J. (1989). Validity of multidimensional scaling in the context of structured conceptualization. *Evaluation and Program Planning, 12*(1), 81-86. doi: 10.1016/0149-7189(89)90026-8

Dunn, J. C. (1974). Well separated clusters and optimal fuzzy partitions. *Cybernetics and Systems: An International Journal 4*(1), 95-104. doi: 10.1080/01969727408546059

Dunn, J. C. (1976). Indicies of partition fuzziness and the detection of clusters in large data sets. In M. Gupta (Ed.), *Fuzzy Automata and Decision Processes*. New York: Elsevier.

Dunn, W. (1981). *Public policy analysis: An introduction*. Englewood Cliffs, NJ: Prentice Hall.

Eisen, M., Zellman, G., & McAlister, A. (1992). A health belief model-social learning theory approach to adolescents' fertility control: Findings from a controlled field trial. *Health Education Quarterly, 19*(2), 249-262.

Everitt, B. S., & Rabe-Hesketh, S. (1997). *The analysis of proximity data*. London: Arnold.

Fitz-Gibbon, C. T., & Morris, L. L. (1996/1975). Theory-based evaluation. *Evaluation Practice, 17*(2), 177-184. doi: 10.1016/S0886-1633(96)90024-0

Flay, B. R., Miller, T. Q., Hedeker, D., Siddiqui, O., Britton, C. F., Brannon, B. R., et al. (1995). The Television, School and Family Smoking Prevention and Cessation Project, VIII. Student outcomes and mediating variables. *Preventive Medicine, 24*(1), 29-40. doi: 10.1006/pmed.1995.1005

Frechtling, J. (2007). *Logic modeling methods in program evaluation*. San Francisco: Jossey-Bass.

Freire, P. (2000). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). New York: Continuum International Publishing Group.

Giurcaneanu, C. D., Tabus, I., Shmulevich, I., & Zhang, W. (2003, July). *Stability-based cluster analysis applied to microarray data*. Paper presented at the Seventh International Symposium on Signal Processing and Its Applications, Paris, France. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1224814&tag=1

Guba, E. G. (Ed.). (1990). *The Paradigm Dialog*. Newbury Park: Sage.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (3rd ed., pp. 191-215). Thousand Oaks, CA: Sage.

Gusfield, D. (2002). Partition-distance: A problem and class of perfect graphs arising in clustering. *Information Processing Letters, 82*(3), 159-164. doi: Doi: 10.1016/s0020-0190(01)00263-0

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (n.d.). Clustering Validity Checking Methods: Part II. Retrieved from http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.4274

Holder, H. D., Saltz, R. F., Treno, A. J., Grube, J. W., & Voas, R. B. (1997). Evaluation design for a community prevention trial. *Evaluation Review, 21*(2), 140-166.

Holton, G. (1993). *Science and anti-science*. Cambridge, MA: Harvard University Press.

Ivankova, N., & Kawamura, Y. (2010). Emerging trends in the utilization of integrated designs in the social, behavioral and health sciences. In C. Teddlie & A. Tashakkori (Eds.), *Sage Handbook of Mixed Methods in Social and Behavioral Research* (pp. 581-611). Thousand Oaks, CA: Sage.

Jackson, K. M., & Trochim, W. M. K. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods, 5*(4), 307-336. doi: 10.1177/109442802237114

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26. doi: 10.3102/0013189x033007014

Kane, M., & Trochim, W. M. K. (2007). *Concept mapping for planning and evaluation*. Thousand Oaks, CA: Sage.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: John Wiley & Sons.

Krippendorf, K. (1980). *Content analysis: An introduction to its methodology* (1st ed.). Beverly Hills, CA: Sage.

Krippendorf, K. (2004). *Content analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *The American Journal of Evaluation, 24*(1), 5-20. doi: 10.1016/S1098-2140(02)00271-0

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1-55.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.

Lincoln, Y. S., & Lynham, S. A. (2011). Criteria for assessing theory in human resource development from an interpretive perspective. *Human Resource Development International, 14*(1), 3-22.

Linton, R. (1989). Conceptualizing feminism: Clarifying social science concepts. *Evaluation and Program Planning, 12*(1), 25-29.

Lipsey, M. W., Crosse, S., Dunkle, J., Pollard, J., & Stobart, G. (1985). Evaluation: The state of the art and the sorry state of the science. *New Directions for Program Evaluation, 1985*(27), 7-28.

Lynham, S. A. (2002a). The general method of theory-building in applied disciplines. *Advances in Developing Human Resources, 4*(3), 221-241. doi: 10.1177/1523422302043002

Lynham, S. A. (2002b). Quantitative research and theory-building: Dubin's method. *Advances in Developing Human Resources, 4*(3), 242-276. doi: 10.1177/15222302004003003

Maechler, M. (2011). Package 'cluster'. Retrieved March 23, 2011, from http://cran.r-project.org/

McKelvey, B. (2002). Model-centered organization science epistemology. In J. A. C. Baum (Ed.), *The Blackwell Companion to Organizations* (pp. 752-780). Oxford: Blackwell.

McNemar, Q. (1969). *Psychological Statistics* (4th ed.). New York: Wiley.

Metro Organizations for People. (n.d.a). *About MOP*. Retrieved August 31, 2010, from the Metro Organizations for People web site: http://www.mopdenver.org/about

Metro Organizations for People. (n.d.b). *Accomplisments and Successes* Retrieved June 10, 2011, from the Metro Organizations for People web site: http://www.mopdenver.org/issues?id=0008

Metro Organizations for People. (n.d.c). Home Page Retrieved June 10, 2011, from the Metro Organizations for People web site: http://www.mopdenver.org

Metro Organizations for People. (n.d.d). *Issues and Results.* Retrieved August 31, 2010, from the Metro Organizations for People web site: http://www.mopdenver.org/issues

Morgan, G. A., Gliner, J. A., & Harmon, R. J. (2006). *Understanding and evaluating research in applied and clinical settings*. Mahwah, NJ: Lawrence Erlbaum Associates.

Murray, D. M., Prokhorov, A. V., & Harty, K. C. (1994). Effects of a statewide antismoking campaign on mass media messages and smoking beliefs. *Preventive Medicine, 23*(1), 54-60. doi: 10.1006/pmed.1994.1008

Nieweglowski, L. (2009). Package 'clv' Retrieved March 25, 2001, from http://cran.r-project.org/

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw Hill.

Osborn, A. F. (1948). *Your creative power*. New York: Scribner.

Patterson, C. H. (1986). Introduction. *Theories of counselling and psychotherapy* (pp. xvii-xxvii). New York: Harper and Row.

Paulson, B. L., & Worth, M. (2002). Counseling for suicide: Client prespectives. *Journal of Counseling and Development, 80*(1), 86-93.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

Petrosino, A. (2000). Whether and why? The potential benefits of including program theory evaluation in meta-analysis. *New Directions for Evaluation, 2000*(87), 59-69. doi: 10.1002/ev.1182

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and Educational Research*. Lanham, MD: Rowman and Littlefield.

PICO National Network. (n.d.). *About PICO*. Retrieved August 31, 2010, from the PICO Web site: http://www.piconetwork.org/about

Puska, P., Nissinen, A., Tuomilehto, J., Salonen, J. T., Koskela, K., McAlister, A., et al. (1985). The community-based strategy to prevent coronary heart disease: Conclusions for the ten years of the North Karelia Projects. *Annual Review of Public Health, 6*, 147-193. doi: 10.1146/annurev.pu.06.050185.001051

Quay, H. C. (1979). The three faces of evaluation: What can be expected to work. In L. Sechrest, S. G. West, M. A. Phillips, R. Radner & W. Keaton (Eds.), *Evaluation studies review annual* (Vol. 4). Beverly Hills, CA: Sage.

R Development Core Team. (2011). R: A language and environment for statistical computing (Version 2.12.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.r-project.org

Ridde, V. (2008). Equity and health policy in Africa: Using concept mapping in Moore (Burkina Faso). *BMC Health Services Research, 8*(90). Retrieved from http://www.biomedcentral.com/1472-6963/8/90 doi:10.1186/1472-6963-8-90

Ripley, B. (2011). Package 'MASS'. Retrieved March 23, 2011, from http://cran.r-project.org/

Rogers, P. J., & Weiss, C. H. (2007). Theory-based evaluation: Reflections ten years on. *New Directions for Evaluation, 2007*(114), 63-81. doi: 10.1002/ev.225

Root, M. (1993). *Philosophy of social science: The methods, ideals, and politics of social inquiry*. Cambridge, MA: Blackwell.

Rosas, S. R. (2005). Concept mapping as a technique for program theory development: An illustration using family support programs. *American Journal of Evaluation, 26*(3), 389-401. doi: 10.1177/1098214005278760

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research, 10*(4), 489-502.

Ross, S. (1988). *A First Course in Probability* (3rd ed.). New York: Macmillan.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton-Mifflin.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I. *Psychometrika, 27*(2), 125-140. doi: 10.1007/bf02289630

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part II. *Psychometrika, 27*(3), 219-246. doi: 10.1007/bf02289621

Snyder, R. (n.d.). *PICO model training*. PICO National Network. Oakland, CA.

Spearman, C. (1910). Correlation calculated with faulty data. *British Journal of Psychology, 3*(3), 271-295. doi: 10.1111/j.2044-8295.1910.tb00206.x

Sridharan, S., Campbell, B., & Zinzow, H. (2006). Developing a stakeholder-driven anticipated timeline of impact for evaluation of social programs. *American Journal of Evaluation, 27*(2), 148-162. doi: 10.1177/1098214006287990

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, MA: The Massachusets Institute of Technology.

Suchman, E. (1967). *Evaluative research*. New York: Russell Sage Foundation.

Taut, S., Santelices, V., Araya, C., & Manzi, J. (2010). Theory underlying a national teacher evaluation program. *Evaluation and Program Planning, 33*(4), 477-486. doi: 10.1016/j.evalprogplan.2010.01.002

Teddlie, C., & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In C. Teddlie & A. Tashakkori (Eds.), *Sage Handbook of Mixed Methods in Social and Behavioral Research* (2nd ed., pp. 1-41). Thousand Oaks, CA: Sage.

Tilley, N. (2000, September). *Realistic evaluation: An overview*. Paper presented at the Founding Conference of the Danish Evaluation Society. Retrieved from http://www.evidence-basedmanagement.com/research_practice/articles/nick_tilley.pdf

Trochim, W. M. K. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning, 12*(1), 1-16. doi: 10.1016/0149-7189(89)90016-5

Trochim, W. M. K. (1993, November). *The reliability of concept mapping*. Paper presented at the American Evaluation Association Annual Conference, Dallas, TX. Retrieved from http://www.socialresearchmethods.net/research/Reliable/reliable.htm

Trochim, W. M. K. (1998). An evaluation of Michael Scriven's minimalist theory: The least theory that practice requires. *American Journal of Evaluation, 19*(2), 243-249. doi: 10.1016/S1098-2140(99)80202-1

Trochim, W. M. K., Cook, J. A., & Setze, R. J. (1994). Using concept mapping to develop a conceptual framework of staff's views of a supported employment program for individuals with severe mental illness. *Journal of Consulting and Clinical Psychology, 62*(4), 766-775. doi: 10.1037/0022-006X.62.4.766

Trochim, W. M. K., & Kane, M. (2005). Concept mapping: an introduction to structured conceptualization in health care. *International Journal for Quality in Health Care, 17*(3), 187-191. doi: 10.1093/intqhc/mzi038

Trochim, W. M. K., & Linton, R. (1986). Conceptualization for evaluation and planning. *Evaluation and Program Planning, 9*(4), 289-308.

Trochim, W. M. K., Marcus, S. E., Masse, L. C., Moser, R. P., & Weld, P. C. (2008). The evaluation of large research initiatives: A participatory integrative mixed-methods approach. *American Journal of Evaluation, 29*(1), 8-28. doi: 10.1177/1098214007309280

University of Wisconsin-Extension. (2002). *Enhancing program performance with logic models*. Retrieved June 7, 2011, from http://www.uwex.edu/ces/pdande/evaluation/pdf/lmcourseall.pdf

Van de Ven, A. H. (2007). *Engaged scholarship: a guide for organizational and social research*. Oxford, England: Oxford University Press.

W. K. Kellog Foundation. (2001). Logic model development guide Retrieved June 7, 2011, from http://www.wkkf.org/knowledge-center/resources/2006/02/WK-Kellogg-Foundation-Logic-Model-Development-Guide.aspx

Warren, M. R. (2001). *Dry bones rattling: community building to revitalize American democracy*. Princeton, NJ: Princeton University Press.

Weiss, C. H. (1972). *Evaluation research: Methods for assessing program effectiveness*. Englewood Cliffs, NJ: Prentice Hall.

Weiss, C. H. (1996/1972). Excerpts from evaluation research: Methods of assessing program effectiveness. *Evaluation Practice, 17*(2), 173-175. doi: 10.1016/S0886-1633(96)90023-9

Weiss, C. H. (1997a). How can theory-based evaluation make greater headway? *Evaluation Review, 21*(4), 501-524.

Weiss, C. H. (1997b). Theory-based evaluation: Past, present, and future. *New Directions for Program Evaluation, 1997*(76), 41-55. doi: 10.1002/ev.1086

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *The Academy of Management Review, 14*(4), 490-495.

Wholey, J. S. (1979). *Evaluation: Promise and performance*. Washington, DC: Urban Institute.

Wood, R. L. (2002). *Faith in action: Religion, race, and democratic organizing in America*. Chicago: University of Chicago Press.

Worthen, B. (1996). Editor's note: The origins of theory-based evaluation. *Evaluation Practice, 17*(2), 169-171. doi: 10.1016/S0886-1633(96)90022-7

Yampolskaya, S., Nesman, T. M., Hernandez, M., & Koch, D. (2004). Using concept mapping to develop a logic model and articulate a program theory: A case example. *American Journal of Evaluation, 25*(2), 191-207.

Appendix A

List of Brainstormed Statements – Lunchtime Session

1. People find their individual voice
2. Some people hesitate about joining the group
3. Some people only listen to see who you are and if you are serious
4. Meet people you have never met before
5. Find out common needs & wants with others
6. People learn leadership skills
7. Gain information & knowledge of processes of gov't
8. Leaders become part of democratic process
9. Relationships are formed
10. Build a sense of community within the large committee
11. Build a sense of community within MOP at large
12. Identify groups and people who are in opposition to our agenda items
13. Solve problems
14. Make systemic change
15. Gain technical skills like computer skills
16. Get better idea of what is going on in community
17. More likely to get involved in other parts of community
18. Gain other perspectives
19. develop skills (analytical, problem solving)
20. develop power and identify power both personally and as a group
21. learn how to fight for an issue whether personally impacted or not because it's the right thing to do
22. increased empathy
23. Link faith and beliefs to action
24. Way to solve problems even if don't speak English
25. Find other issues that are linked to your issue for example education and healthcare are connected
26. Parent won getting classrooms as colleges at Bruce Randolph
27. Get to know other parents to better children's education and increase communication
28. MOP established Parent Liaisons in schools, PL's help connect us to resources in the school
29. Kids begin asking questions about parent's involvement in organizing work – this is good
30. Challenge for parents to do this work and balance role as parent, if at a meeting who is taking care of child

31. If both spouses aren't involved at the same level there can be tension created by attending so many meetings
32. Mop provides childcare and food so creates pathway for parents to get involved
33. The healthcare committee was challenged to keep going after SCHIP vetoed twice
34. SCHIP signed by President
35. Healthcare legislation – was passed but got ugly
36. We followed the PICO principle "All organizing is reorganizing" to keep the momentum going?
37. Healthcare passed because of relationships MOP built for example with Senator Bennet and Rep DeGette
38. Instrumental in getting first school based clinic opened in Aurora public schools & now serves three public & one Catholic schools and soon will open a 2nd clinic
39. We learned how to be strategic and identify where we have power to impact healthcare bill
40. Identified specific issues like affordability that must be included in final healthcare bill
41. MOP uses data and personal stories to humanize the data to build the case for our agenda
42. As an organization we have group support in our numbers when we hit barriers to support each other
43. We draw on our faith/values to keep going when we hit barriers
44. Name recognition with members of school board, city council, state legislature, congress – with staff members and elected official
45. Nationally we have relationships with high profile partners and stakeholders ex Kaiser, nurses assoc
46. Locally and statewide we have relationships with partners and stakeholders ex heaa
47. Sometimes we have to sacrifice our own personality and style for the larger group
48. With Sun Valley Coalition, residents bring their issues to the committee and the committee can help shed light on those issues and with the groups with power like decator place – the committee is seen as a resource for power to community members
49. We learn to compromise
50. Leaders have to push back on staff, if staff are leading the charge in a different direction or moving forward on a different issue cut
51. Sometimes leaders and staff have to compromise on issue cut or strategy

52. We identified a larger umbrella solution to address a lot of smaller issues in the community (recreation center)
53. Research problems
54. Won the island and so kids are safer but had to push power players to truthfully research the issue
55. Committee pushed the city for traffic study and learned to persevere the solution identified by the city was different than what the committee hoped for – a compromise
56. Neighbors see leaders as a resource information guide especially on the bad 3 People to leaders for information or leadership on
57. When leaders publically speak out on an issue, the leader can sometimes come under attack – example when Susan testified before Congress on SCHIP
58. In conservative environment, leaders risk being labeled liberal if they participate in organizing work and the organizing can alienate some parts of the community
59. Able to educate others about issues and help to shape their views ex. Hold classes or forums on issues such as healthcare
60. Learn so many new skills as a leader: writing press releases, speeches
61. Learn new skills when become a board member to understand how the organization works
62. Participate in non-partisan electoral work including doorknocking, phonebanks
63. Empower others including professionals like reporters, Tina Griego, to improve their work and in the case of Griego how she reported on Sun Valley
64. Develop relationships with media including reporters
65. Develop skills to put an initiative on the ballot
66. Develop skills to figure out what is required to construct a new recreation center
67. Turn out large numbers of people to demonstrations, actions, press conferences
68. Hold research actions with decision makers – learn from them and exert influence over them
69. Leaders learn or have learned to have discussions around divisive issues and deal with conflict constructively

Appendix B

List of Brainstormed Statements – Evening Session

1. Brings people together
2. Transforms people
3. Informs people about issues
4. Personally – brings out best, empowers
5. Shows you your own faults and challenges and how to overcome them
6. Learning to work with people who have different views, viewpoints, beliefs
7. Makes it harder for people to shut me up
8. Fosters understanding of community, who is community – church, neighborhood, other groups around church
9. Helps me to get to know people with whom I would otherwise not have related to or known.
10. How to analyze, understand and solve issues in the community
11. Getting to know myself personally -
12. learn how to interact with someone who has more power,
13. people learn more compassion for adversaries
14. Learning to accept that there are people who will always be outside your sphere of influence – and when to move on.
15. C.O. provides solutions to problems previously thought to be unwinnable.
16. Brings out hope @ Harrington school – wanted parking lot, unsafe for kids – mop helped up get one way street., crossing, bigger parking lot and getting a new parking lot. Helped our parents see that there was hope.
17. When MOP organizes the city officials listened more – to parents
18. MOP showed people how to have a voice
19. Learning how political system works.
20. Learning about public officials – that they do want to help, not just them vs us, rather partnering with us.
21. Learning why voting matters
22. When mop organizes small groups come together to build more power.
23. PICO model doesn't just fix community helps me know what my values are, how they reflect in my life.
24. Helps me analyze myself and what I've learned.
25. How gov't and systems work – pulled back layers of onion – why such and such statute or law is the way it is.
26. When mop organizes people's needs are clarified
27. Come to realize the needs / self – interests of politicians as well
28. MOP helped get a health clinic in Commerce city – Salud clinic – now it's own building – soon 10 more doctors – people get jobs as well

29. Find out city council is happy that people show up and talk with them. City officials sometimes love that they are being addressed by community
30. We won housing relocation assistance for mobile home park tenants – people got stipends
31. Leaders are born, present themselves in different ways and become more confident and empowered
32. A leader, who never spoke in public, now helps run a meeting; When leaders see that people will listen – it changes them.
33. Made me feel that I do have a voice, that I am an intelligent person that what I have to say is important even though I don't have a degree.
34. Convert cheerleaders into coaches – Listeners become doers
35. You learn how to recruit people
36. How to teach people how to see their self interest in a common goal
37. Learn how to listen people's stories and identify their issues
38. How to tell your own story
39. How to get our story heard in the broader community
40. Puts a face on issues, humanizes problems in the community
41. Help public officials develop better sense of responsibility and accountability
42. We become educated on how systems works, organizing, now to talk, how to get your point across – education, health care, immigration, jobs, banks.
43. How to come to a group of solutions – multiple solutions to an issue.
44. Teaches us how to do research
45. MOP opened my eyes to issues that I had never paid attention to – for ex immigration / DREAM act
46. When MOP organizes, people get out the vote – defeated the bad 3
47. Learn about the data re: how our schools are performing –
48. Education wins – new transportation network in NNE (near ne Denver)
49. Education – won Weighted student funding – in DPS
50. When mop organizes my principals asks to work more with mop and other principals – working together in NNE Denver.
51. When mop organizes within Faith communities it challenges us to examine our morality and spirituality and how it can / could inform how we vote on issues.
52. Helps me to realize that it's not just in my self interest to care about myself
53. Helps us to work to propose a recreation center vote in aurora that will pay for multiple rec. centers in our community.
54. MOP/PICO helped us win SCHIP reauthorization (national); hospital provider fee bill (100,000 new insured). National Health Care reform.
55. Increase library hours in low income neighborhoods.
56. Skills we learn in mop inform other areas in our lives: better, active and engaged community members in my job, community, church, school, life.

Appendix C

Combined List of Brainstormed Statements Used for the Study

1. At Harrington School, MOP helped us get a one way street, a crossing, a bigger parking lot and (soon) a new parking lot. Parents now see there is hope.
2. Healthcare legislation got passed, but it got "ugly" in the process.
3. Healthcare legislation passed because of relationships MOP built; for example, with Senator Bennet and Rep. DeGette.
4. I feel that I do have a voice, I am intelligent and what I have to say is important…even though I don't have a college degree.
5. I get to know people I would otherwise not have known.
6. I learn my own faults & challenges and I learn how to overcome them.
7. I learn to accept that there are people who will always be outside my sphere of influence – and when to move on.
8. I realize that it's not in my self-interest to care just about myself.
9. If we become board members, we learn new skills and we understand how the organization works.
10. In a conservative environment, leaders risk being labeled "liberal" if they participate in organizing work. The organizing can alienate some parts of the community.
11. It is now harder for people to "shut me up".
12. Leaders have to "push back" on staff if the staff is moving forward on a different issue that does not meet needs of leaders.
13. Leaders learn or have learned to have discussions around divisive issues and deal with conflict constructively.
14. MOP established Parent Liaisons in schools; PLs help connect us to resources in the school.
15. MOP had a relationship with Senator Bennett because of the work we did while he was superintendent of DPS.
16. MOP has name recognition with members of the school board, city council, state legislature, congress and with their staff members.
17. MOP has relationships with high profile partners and stakeholders, for example, Kaiser, nurses' association & HEAA.
18. MOP helped get Salud health clinic in Commerce City. It now has its own building and soon 10 more doctors.
19. MOP leaders become part of democratic process.
20. MOP opened my eyes to issues that I never paid attention to, for example, immigration & the DREAM act.
21. MOP puts a "face" on issues and data, humanizes problems in the community.

22. MOP/PICO helped us win the hospital provider fee bill (100,000 new insured).
23. My child's school principal asks to work more with MOP and with other principals in NNE Denver.
24. My kids began asking questions about my involvement in organizing work (this is good).
25. Neighbors viewed MOP leaders as a voting information resource, especially on the "bad 3" propositions in 2010 (60, 61 and 101).
26. One leader, who never spoke in public, now helps run a meeting. When leaders see that people will listen to them, it changes them.
27. Organizing fosters understanding of who is our community – church, neighborhood, groups near the church.
28. Our church committee pushed the city for a traffic study. We learned to persevere when the solution identified by the city was different than what we hoped for. We compromised.
29. Parents won getting classrooms to be like "colleges" at Bruce Randolph School.
30. People got out the vote and defeated the "bad 3" propositions in 2010.
31. Salud clinic provides people jobs.
32. SCHIP was signed by President.
33. Small groups come together to build more power.
34. Some people hesitate to join MOP & only listen to see who you are and if you are serious.
35. The healthcare committee was challenged to "keep going" after SCHIP vetoed twice.
36. The library hours in low-income neighborhoods increased.
37. The PICO organizing model doesn't just fix the community, it helps me know what my values are.
38. The skills and involvement we learn with MOP informs other areas in our lives. We are active and engaged community members and better on the job and at church and in school.
39. The Sun Valley Coalition committee is seen as a resource for power for residents at Decator Place. Residents come to committee members when they need an issue addressed by Decator Place managers.
40. There is a new transportation network benefiting students in NNE Denver.
41. We are able to educate others about issues and help to shape their views. Example: classes or forums on healthcare.
42. We are challenged as parents to balance doing organizing work and taking care of our children. Example: going to a MOP meeting.
43. We are personally empowered and become confident leaders.
44. We become educated on how to talk, how to get our point across and how to get our story heard in the broader community.

45. We build a sense of community within LOCs and with all of MOP and we support each other when there are barriers in the organizing work.
46. We can empower others to improve their work and report more accurately on the Sun Valley Coalition. Example: professional reporter Tina Griego.
47. We can turn out large numbers of people to demonstrations, actions and press conferences.
48. We come to realize the needs & self-interests of politicians.
49. We develop analytical and problem-solving skills.
50. We develop relationships with the media, including reporters.
51. We develop skills to put an initiative on the ballot, for example, Aurora recreation center.
52. We draw on our faith & values to keep going when we hit barriers in organizing.
53. I find that if both spouses aren't involved in organizing at the same level, there can be tension created by attending so many meetings.
54. We find common needs with others and we can identify our self-interest in a common goal.
55. We find solutions to problems previously thought to be unwinnable.
56. We find that issues are linked. Example: education and healthcare are connected.
57. We found out that city council is happy when community members show up and talk with them.
58. We gain technical skills (i.e. computer skills).
59. We get informed about community issues.
60. We get to know other parents so we can better our children's education and increase communication.
61. We help public officials develop a better sense of responsibility and accountability to the people. Example: city officials listening to parents.
62. We identify groups and people who are in opposition to our agenda.
63. We increase empathy and compassion, even for adversaries.
64. We learn how the political process and system works.
65. We learn how to do research on community problems.
66. We learn how to fight for an issue even if we are not personally impacted…because it's the right thing to do.
67. We learn how to interact with someone who has more power and how to learn from and influence them. Example: decision-makers.
68. We learn how to listen to people's stories, clarify their needs and identify their issues.
69. We learn how to recruit other people to participate.
70. We learn skills like writing press releases and speeches.
71. We learn to compromise (including with MOP staff) regarding an issue cut or strategy.

72. We learn to develop power and identify power, both personally and as a group.
73. We learn why voting matters.
74. We learned about data on how our schools are performing.
75. We learned about the process of constructing a new recreation center. Example: zoning laws.
76. We learned that there are multiple solutions to an issue and that one solution can sometimes address many smaller issues. Example: recreation center.
77. We learned to be strategic and to identify where we have power to impact the healthcare bill. We identified specific issues like affordability that had to be included in the final bill.
78. We make systemic change.
79. We meet and learn to work with people who have different viewpoints & beliefs.
80. We participate in non-partisan electoral work including doorknocking and phonebanks.
81. We see that MOP provides childcare and food at meetings which creates a pathway for parents to get involved.
82. We solve problems (even if we don't speak English).
83. We were instrumental in getting the first school-based clinic opened in Aurora public schools; it now serves three public & one Catholic school and soon will open a 2nd clinic.
84. We won an island in the road near the school and kids are safer. We had to push the power players to research this issue truthfully.
85. We won housing relocation assistance for mobile home park tenants – people got stipends.
86. Weighted student funding was implemented in DPS.
87. When MOP leaders publically speak out on an issue, we sometimes come under attack. Example: Susan testifying before Congress on SCHIP.
88. When we had difficulties, we followed the PICO principle "All organizing is reorganizing" to keep the momentum going.
89. Within faith communities, MOP's organizing challenges us to examine our morality and spirituality and how it can inform our vote.

Appendix D

Packet Instructions

ENVELOPE CONTENTS

"Using Concept Mapping as a Tool for Program Theory Development"
AKA: How can we show what happens when MOP organizes?

This envelope should contain the following:

1. 89 numbered statement cards
2. 20 rubber bands
3. Sorting activity sheets
4. Rating activity sheets

Please do the following:

1. First, do the sorting activity and record your answers.
2. Second, do the rating activity.
3. Third, put the following items back in the addressed manila envelope:
    a. All cards, rubber-banded into groups
    b. Sorting activity recording sheets
    c. Rating sheet
4. Seal the envelope and put it in the mail (no postage needed) before 2/25/2011.

If you have any questions at all, please contact Becky Orsi at 970-491-3167 or send me an e-mail at becky.orsi@colostate.edu. If for some reason you are not able to reach me, please contact Kristee Paschall at MOP and she can forward your question to me.

Thanks for your participation in this research project!

Appendix E

Rating Instructions

RATING ACTIVITY

"Using Concept Mapping as a Tool for Program Theory Development"
AKA: How can we show what happens when MOP organizes?

Please complete the SORTING ACTIVITY before you complete the rating activity.

Name:

Please tell me whether each statement listed below occurred in the **short-term** (took less than six months of organizing) or in the **long-term** (took more than six months of organizing).

Statement	Timeframe	
	Short-Term	Long-Term
1. At Harrington School, MOP helped us get a one way street, a crossing, a bigger parking lot and (soon) a new parking lot. Parents now see there is hope.	ST	LT
2. Healthcare legislation got passed, but it got "ugly" in the process.	ST	LT
3. Healthcare legislation passed because of relationships MOP built; for example, with Senator Bennet and Rep. DeGette.	ST	LT
4. I feel that I do have a voice, I am intelligent and what I have to say is important…even though I don't have a college degree.	ST	LT
5. I get to know people I would otherwise not have known.	ST	LT

Appendix F

Demographic Survey

PARTICIPANT BACKGROUND INFORMATION

"Using Concept Mapping as a Tool for Program Theory Development"

Name: _____

1. How many years have you been a leader on a MOP Local Organizing Committee (LOC)?

○	○	○	○	○
Less than 1 year	1 year	2 years	3 years	4 or more years

2. Have you ever been a leader on an LOC with any other PICO organization?

○	○
Yes	No

3. Is your Local Organizing Committee (LOC) affiliated with a faith community?

○	○
Yes	No

4. Is your Local Organizing Committee (LOC) affiliated with a school?

○	○
Yes	No

5. What is your highest level of education?

○	○	○	○
No high school diploma	High school diploma	2-year associates degree	4-year college degree

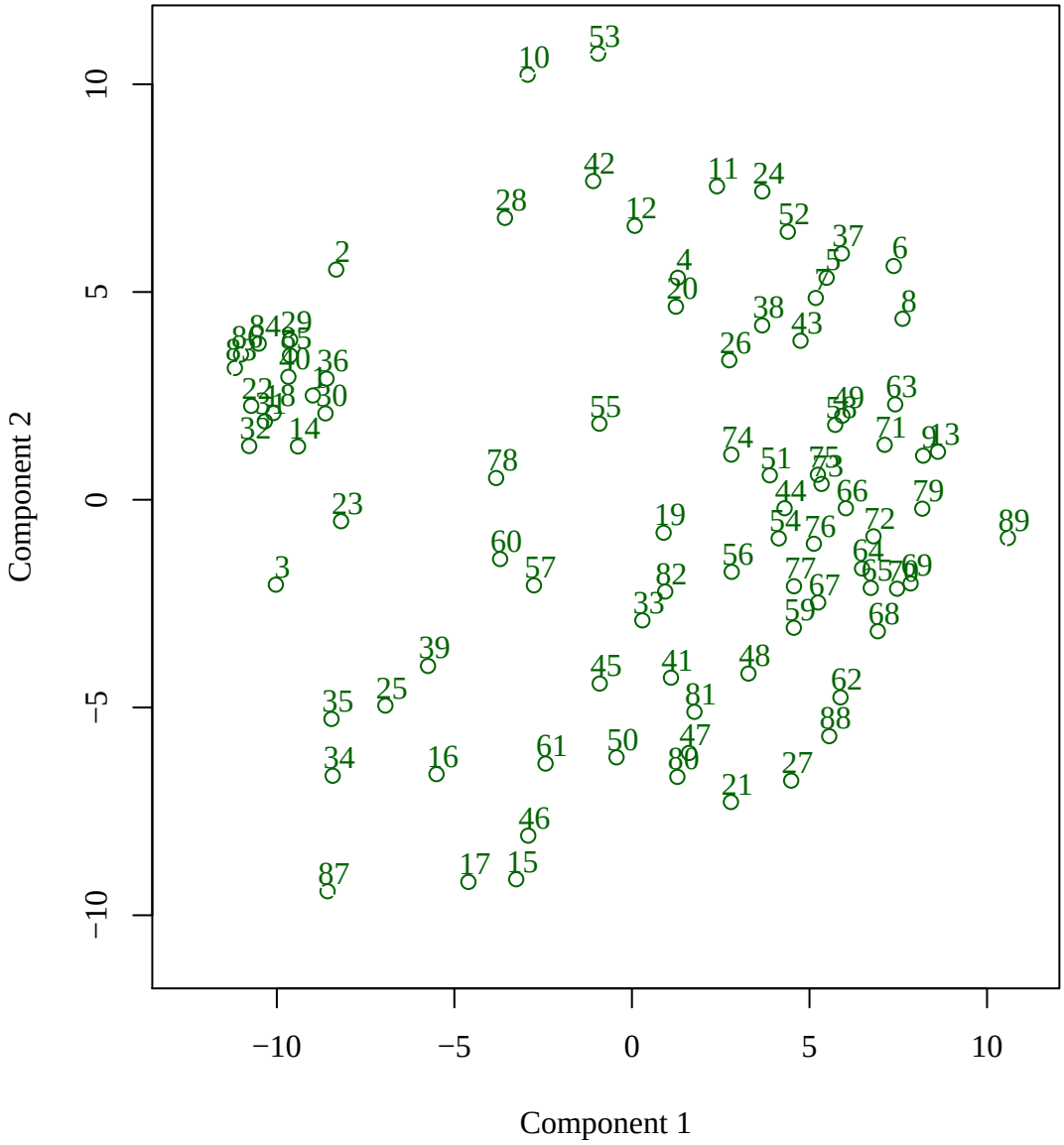
6. What is your age?

7. Do you care for any children under age 18 who live in your home?

○	○
Yes	No

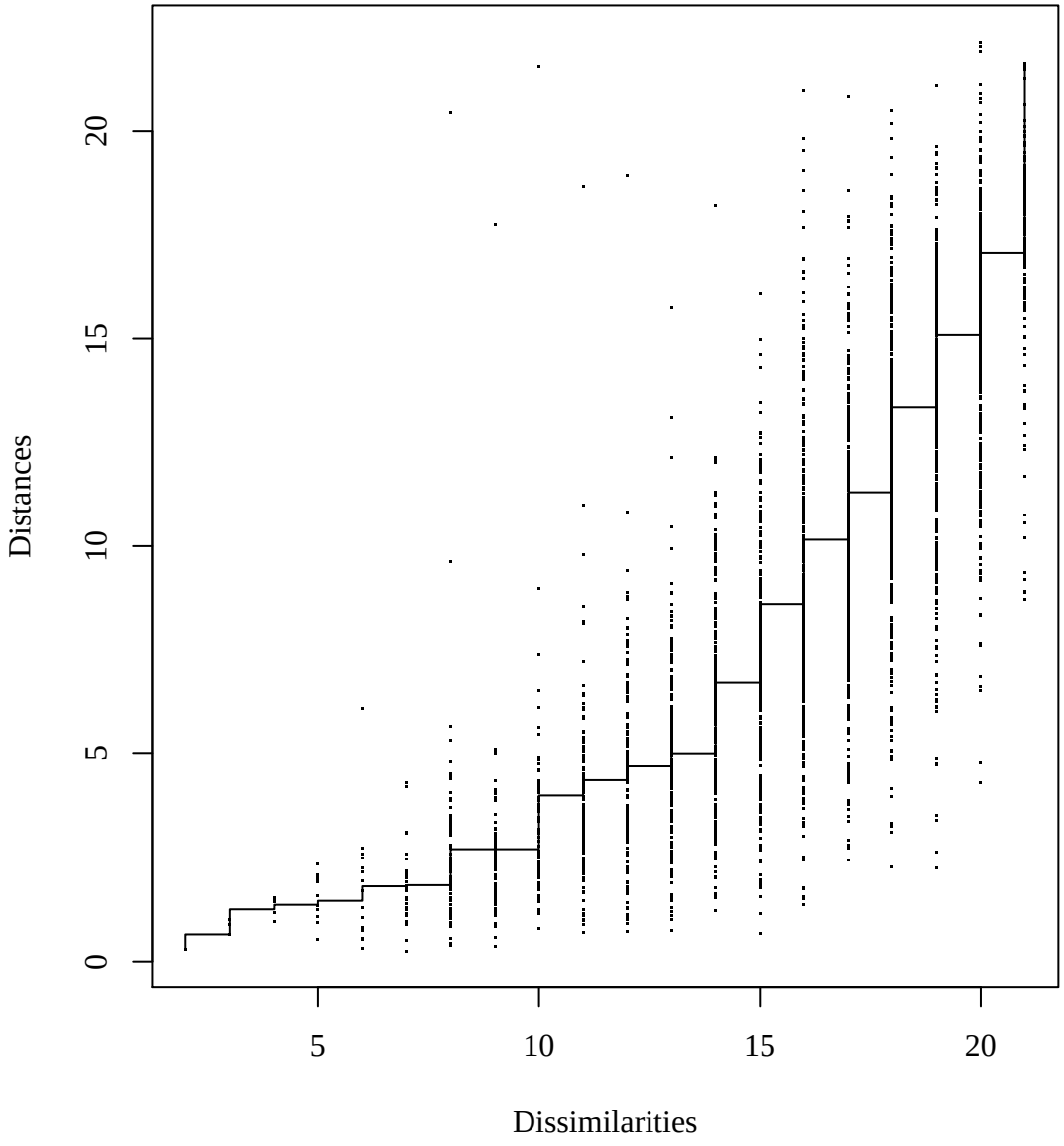
Appendix G

MOP Point Map



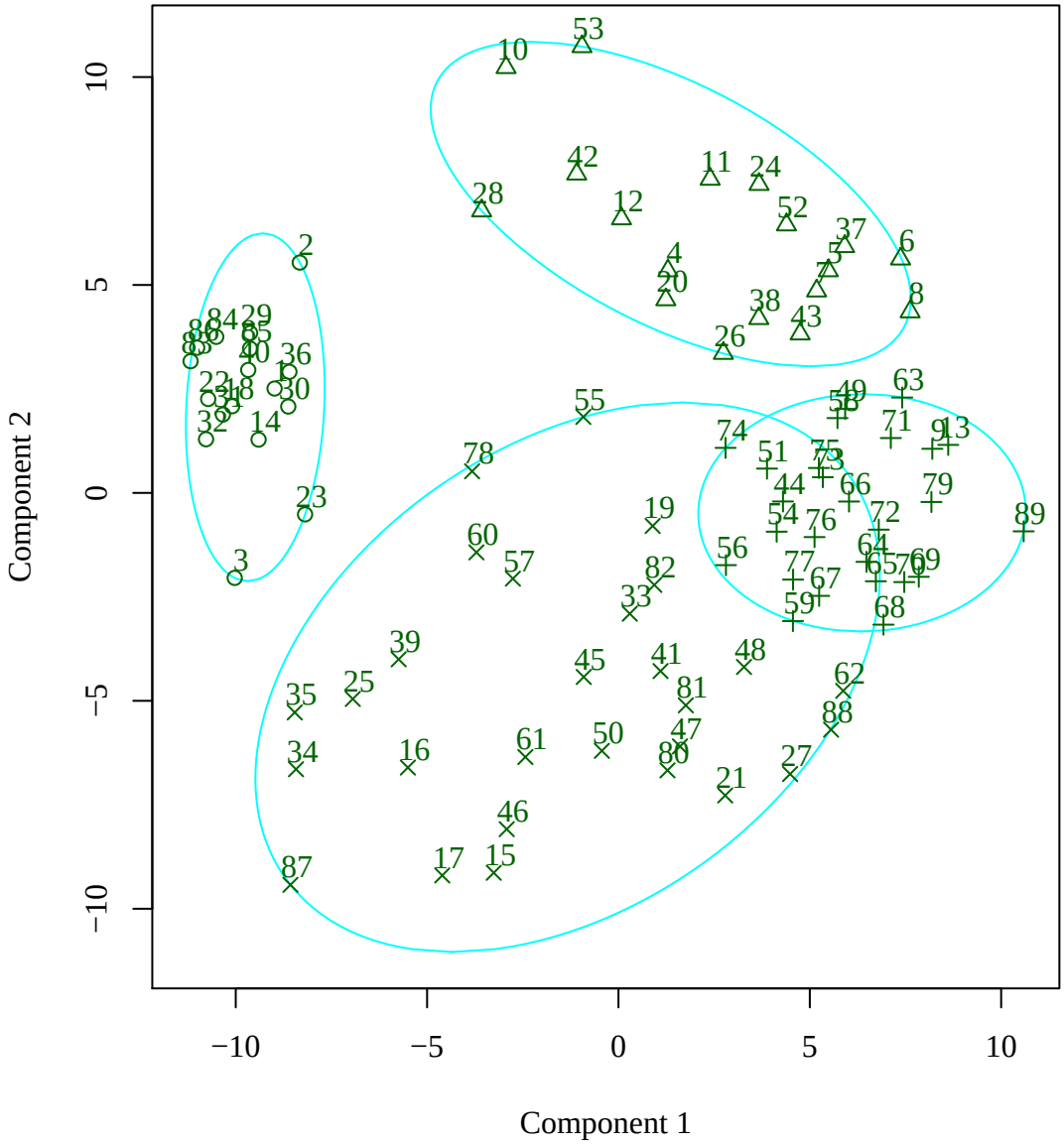
Appendix H

Shepard Diagram



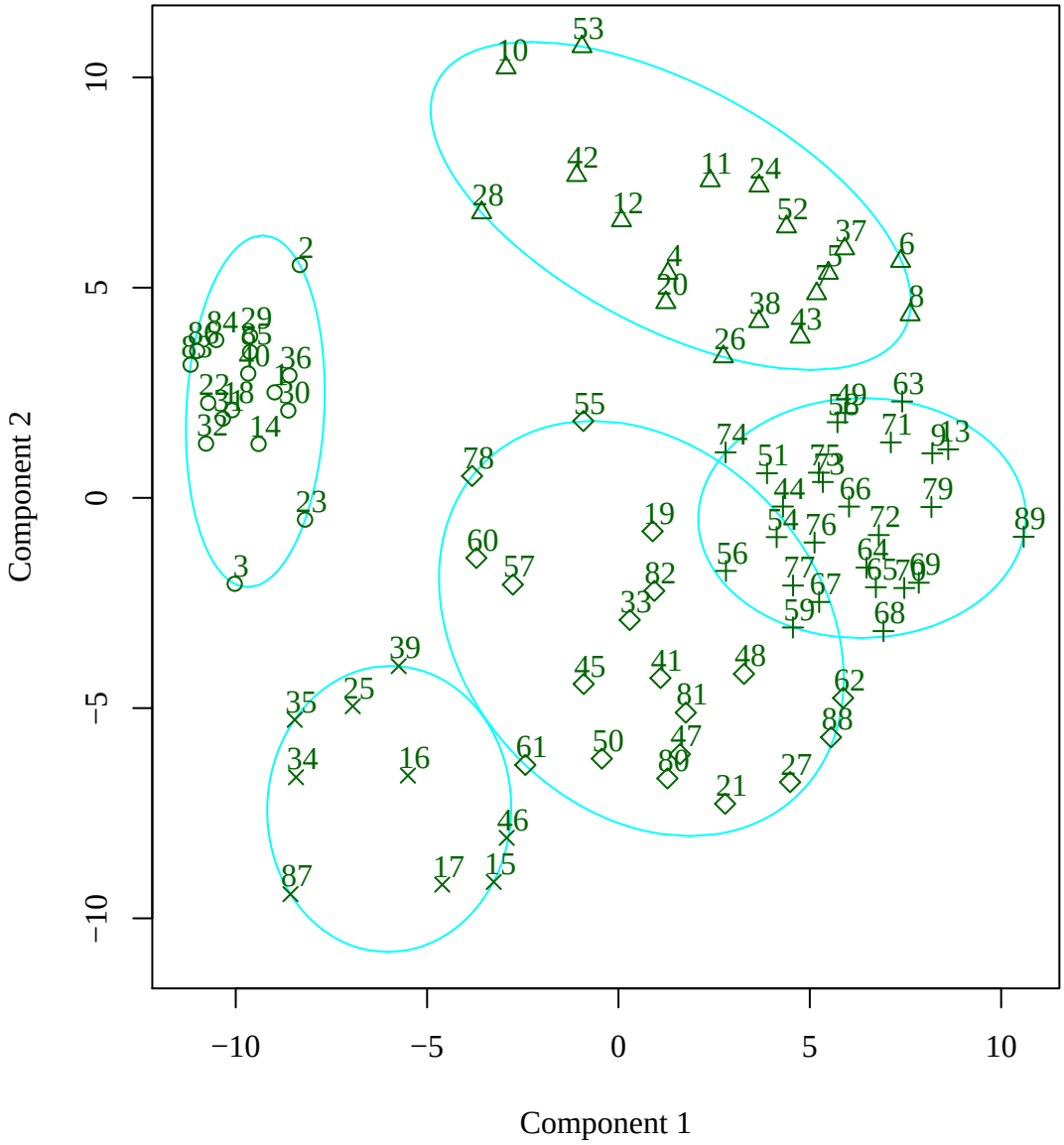
Appendix I

AGNES 4-Cluster Solution



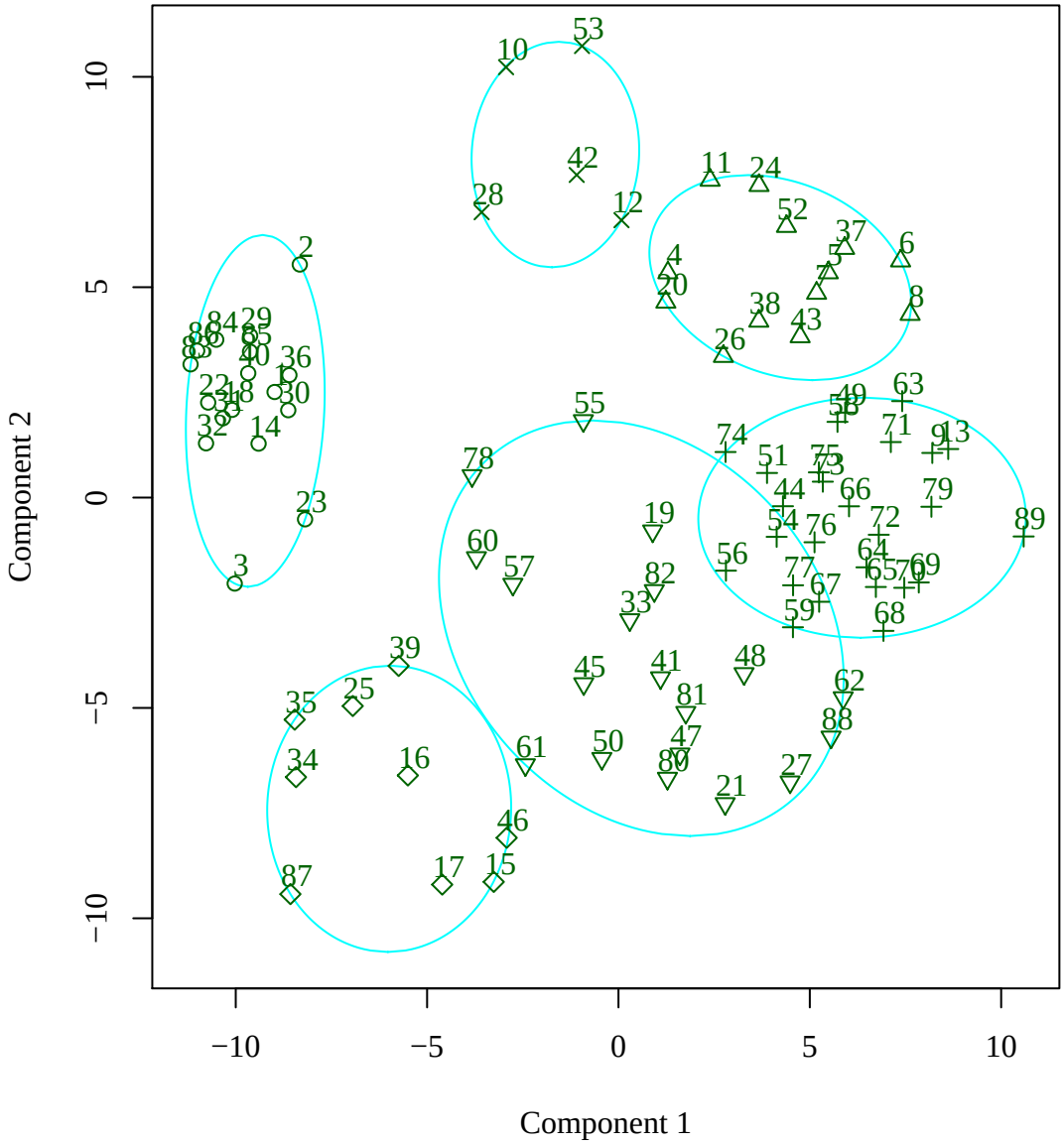
Appendix I

AGNES 5-Cluster Solution



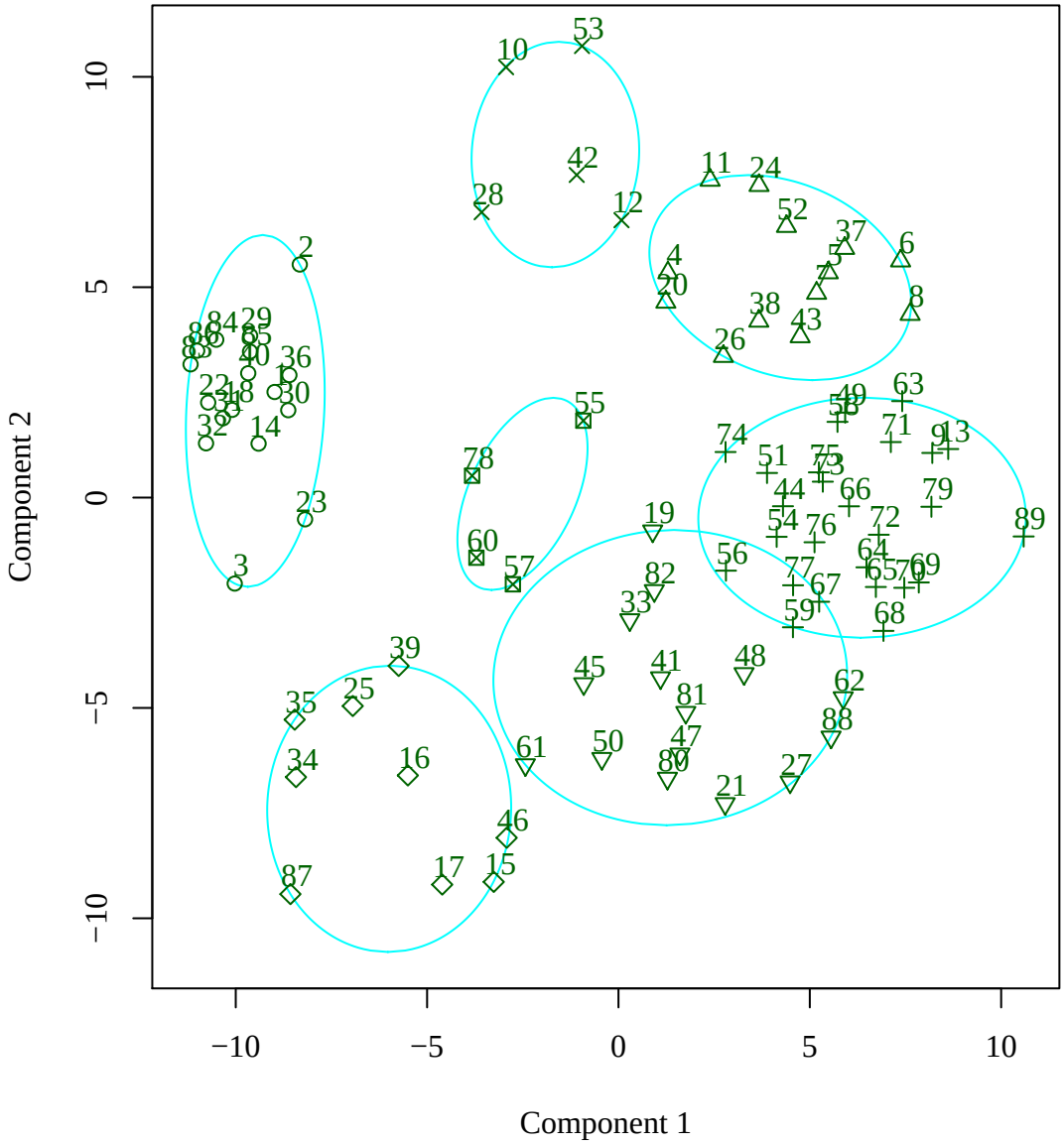
Appendix I

AGNES 6-Cluster Solution



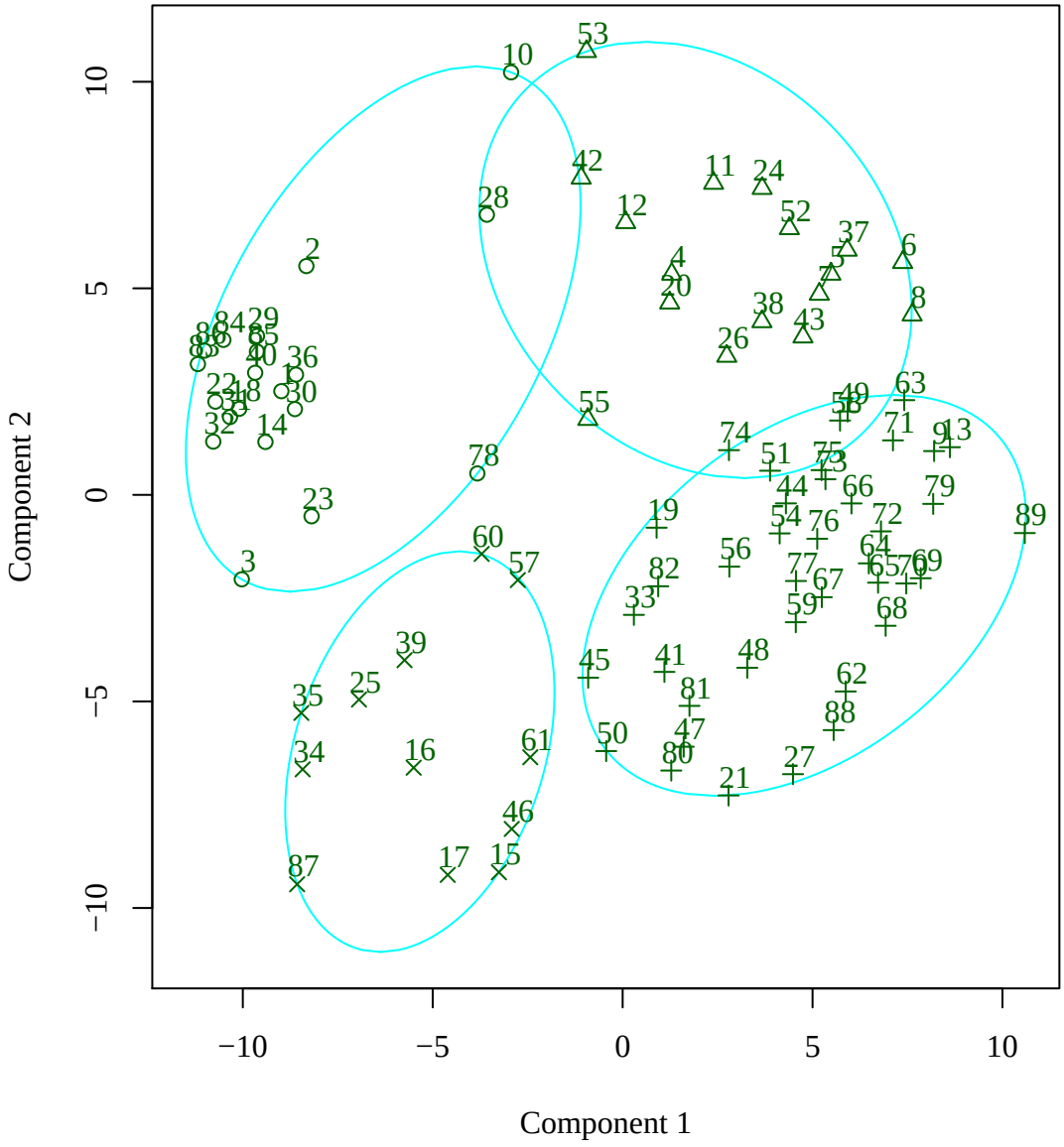
Appendix I

AGNES 7-Cluster Solution



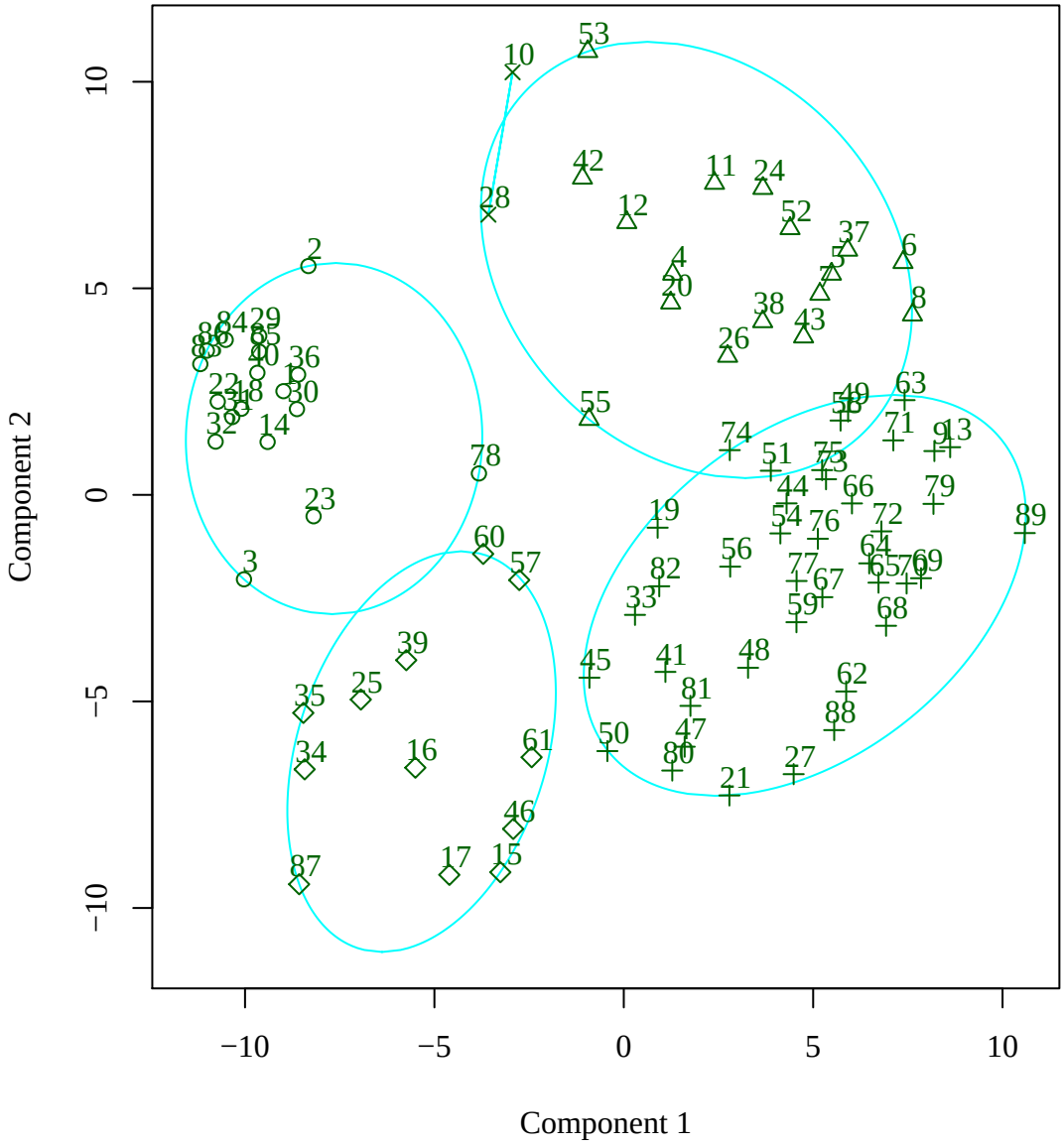
Appendix J

DIANA 4-Cluster Solution



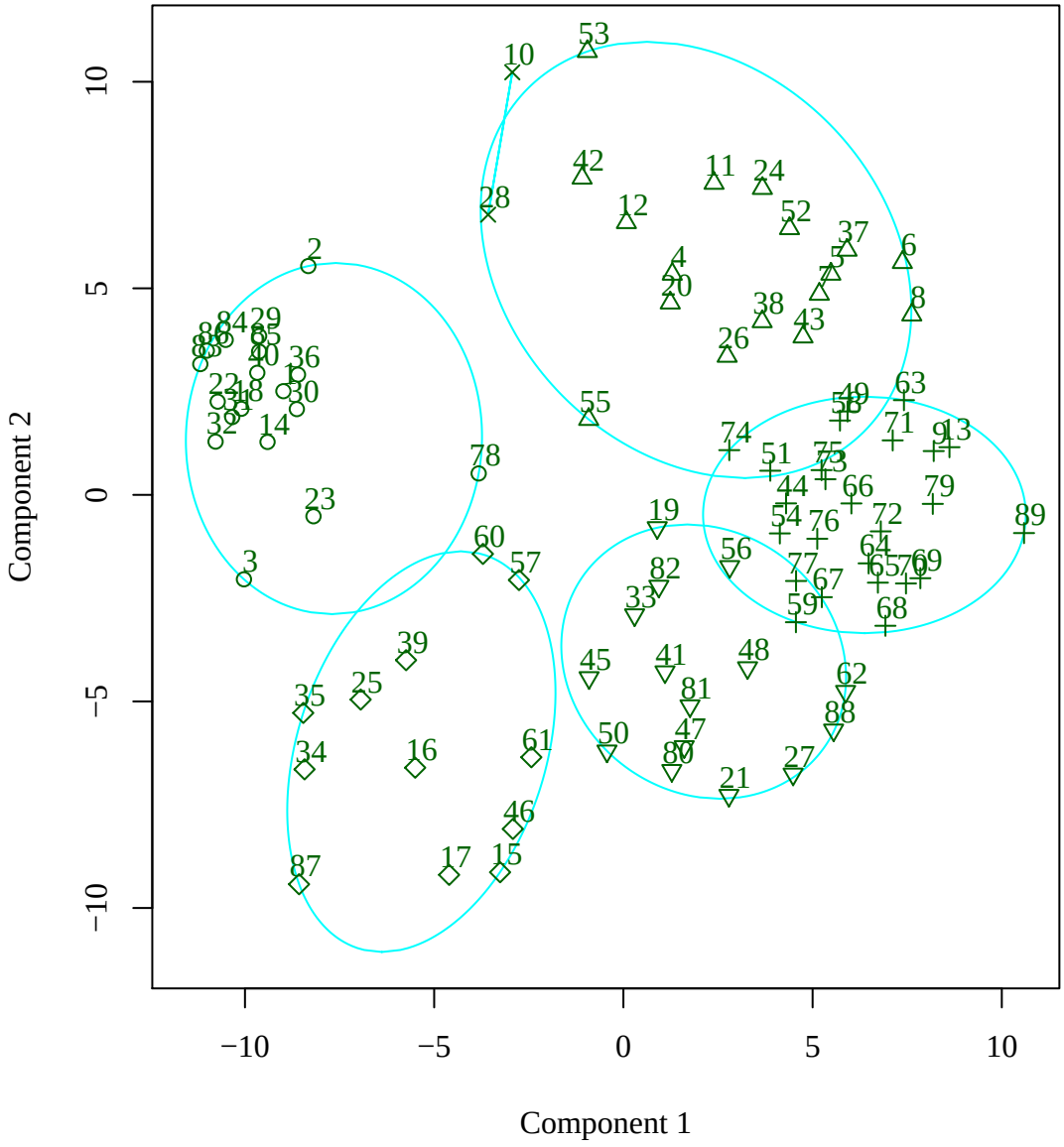
Appendix J

DIANA 5-Cluster Solution



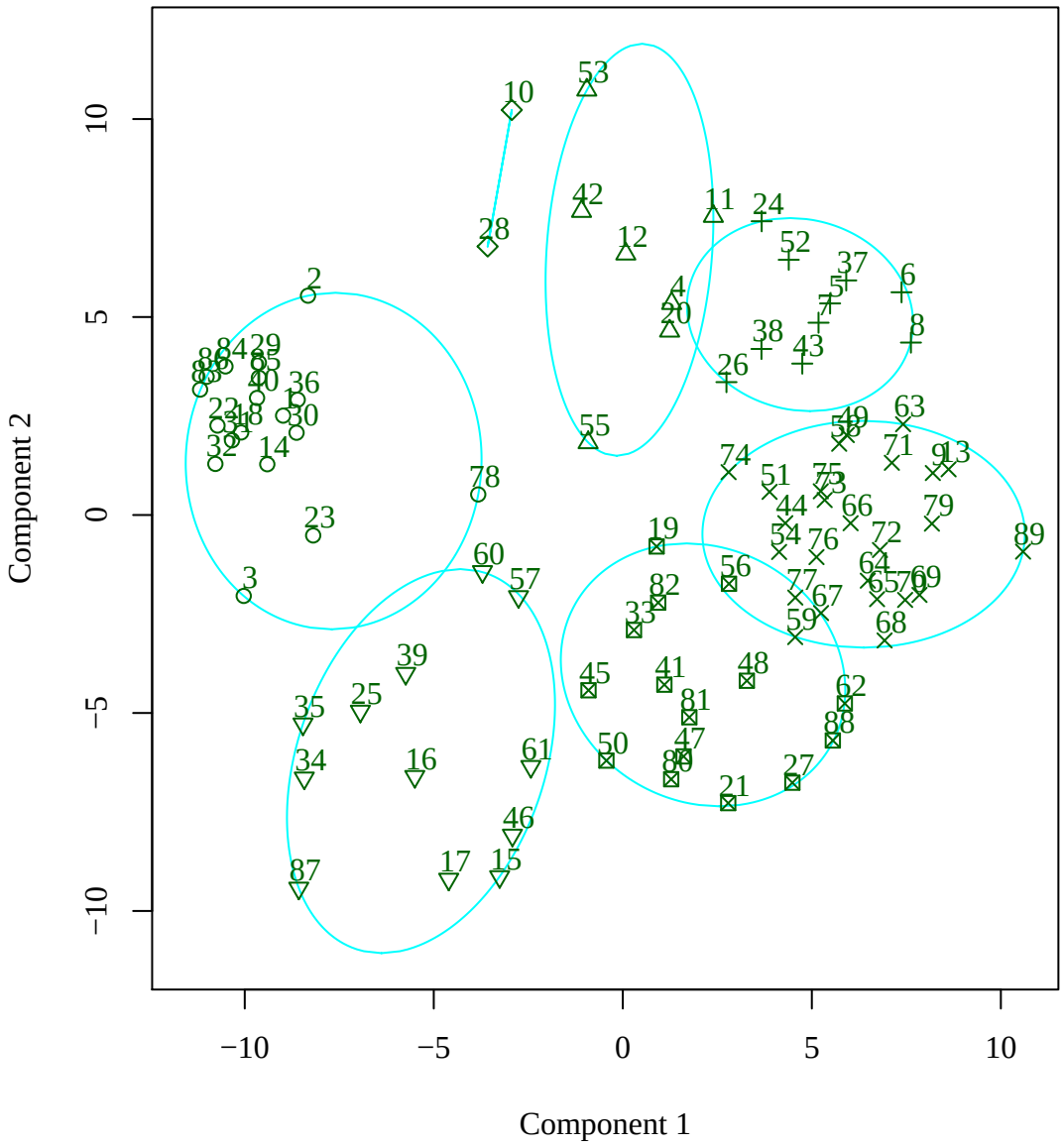
Appendix J

DIANA 6-Cluster Solution



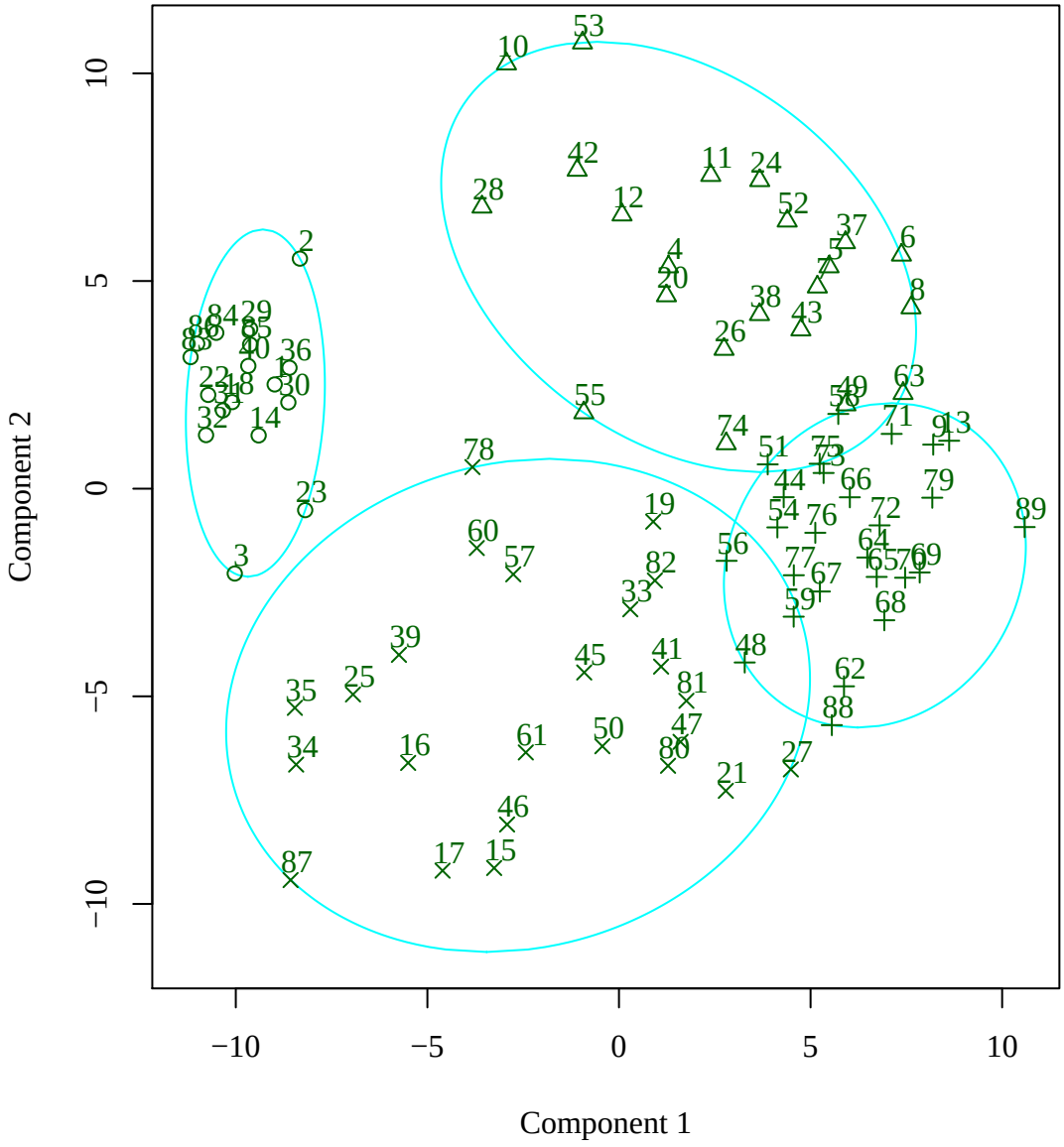
Appendix J

DIANA 7-Cluster Solution



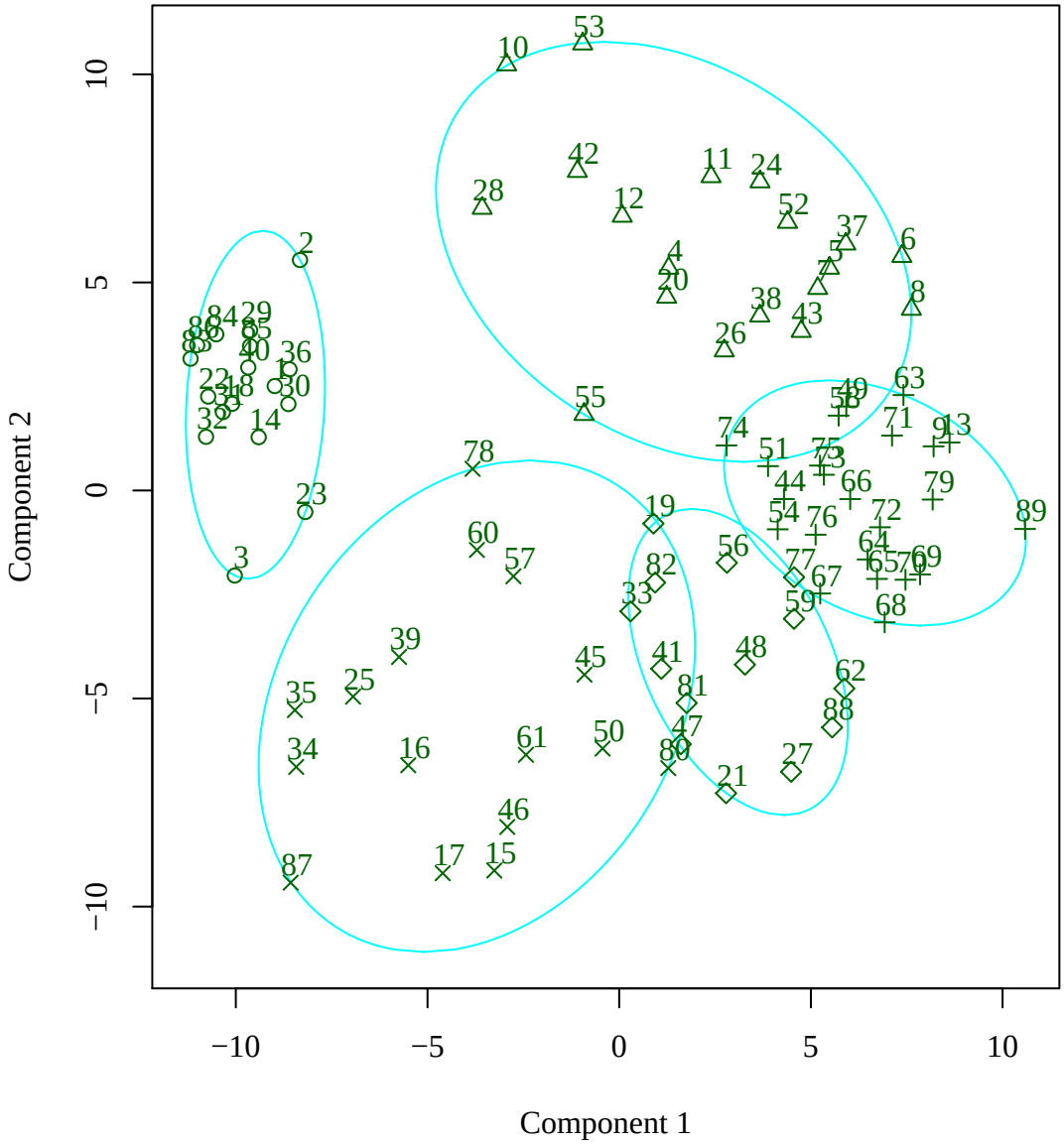
Appendix K

FANNY 4-Cluster Solution



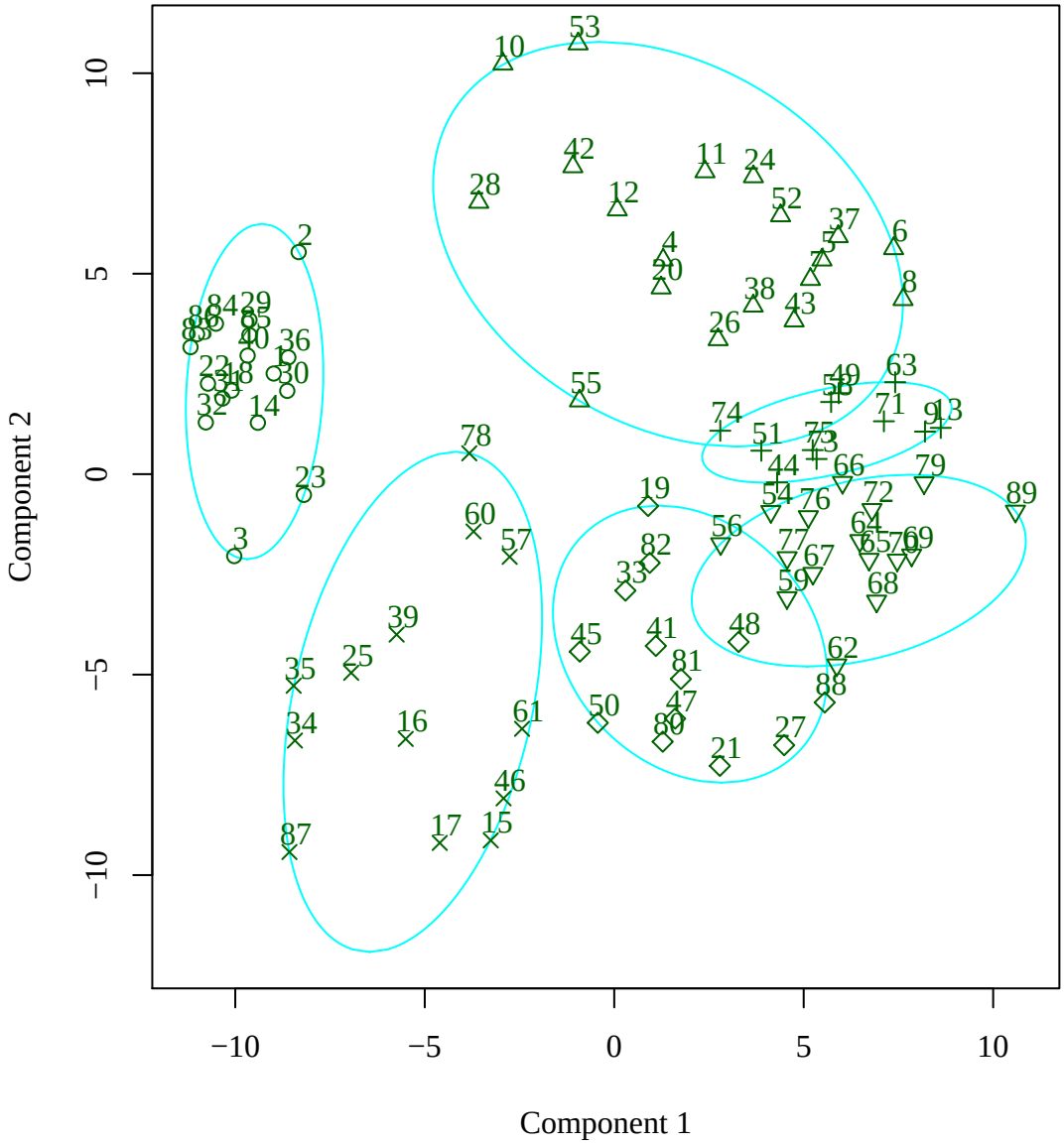
Appendix K

FANNY 5-Cluster Solution



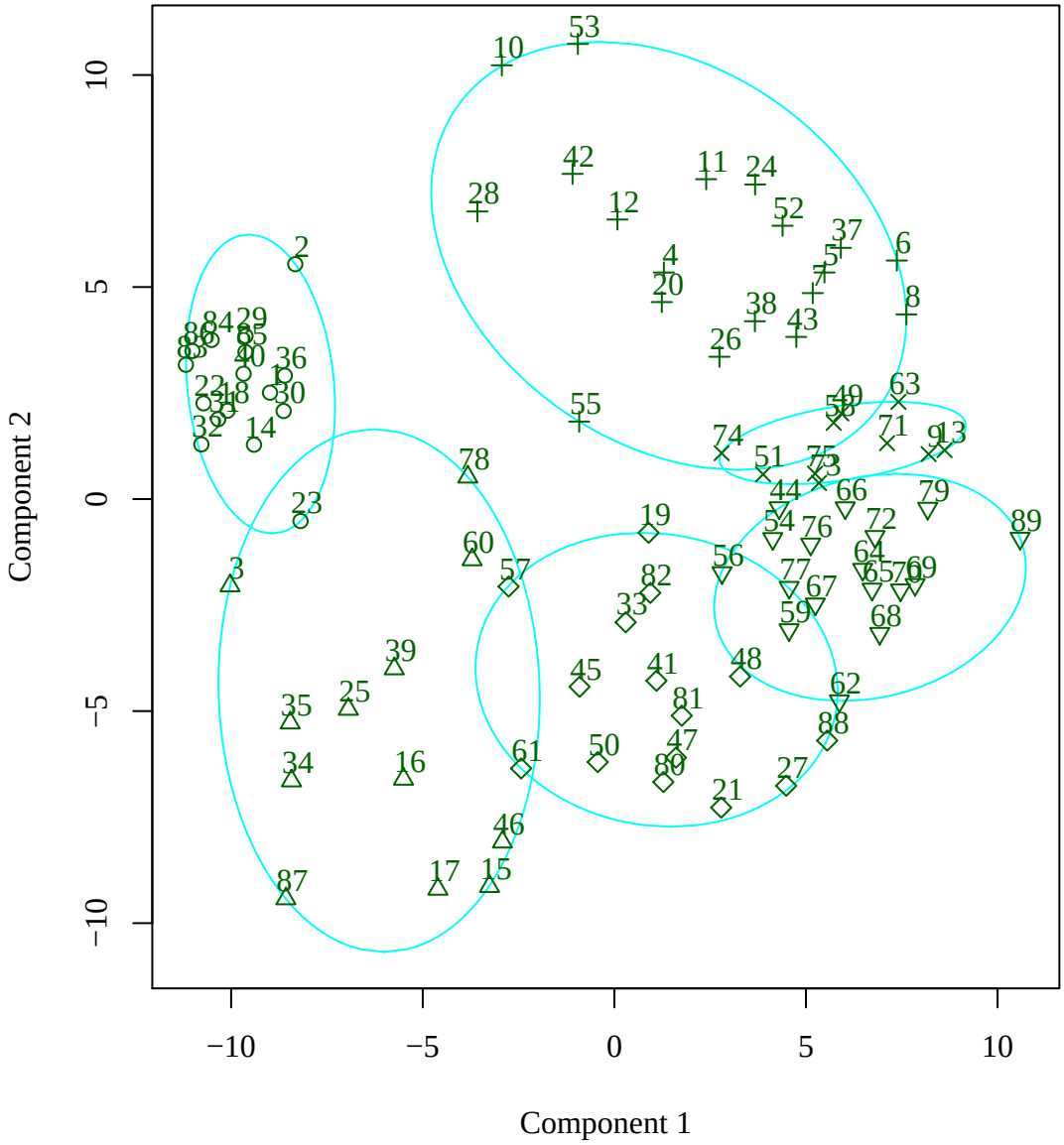
Appendix K

FANNY 6-Cluster Solution



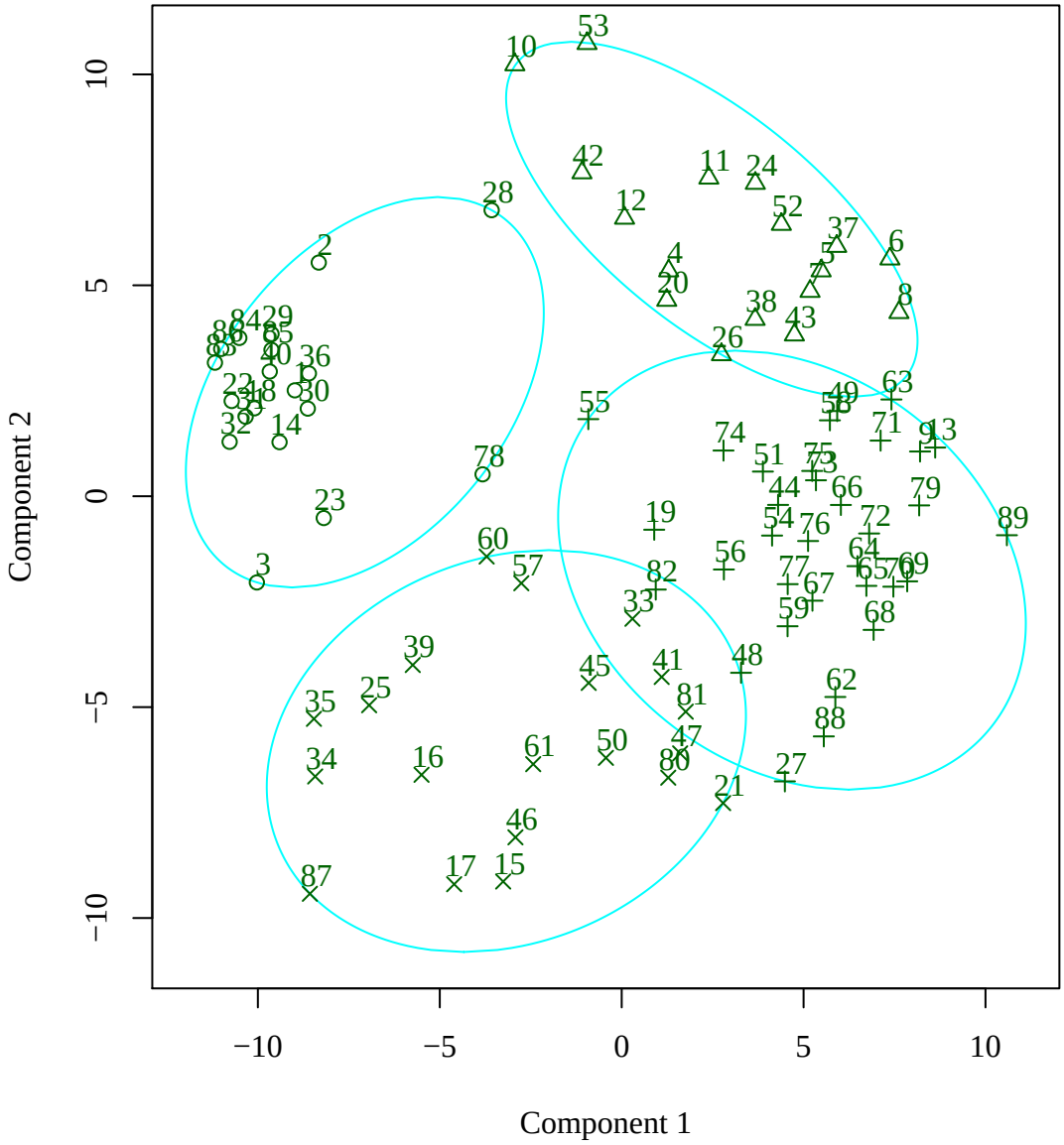
Appendix K

FANNY 7-Cluster Solution



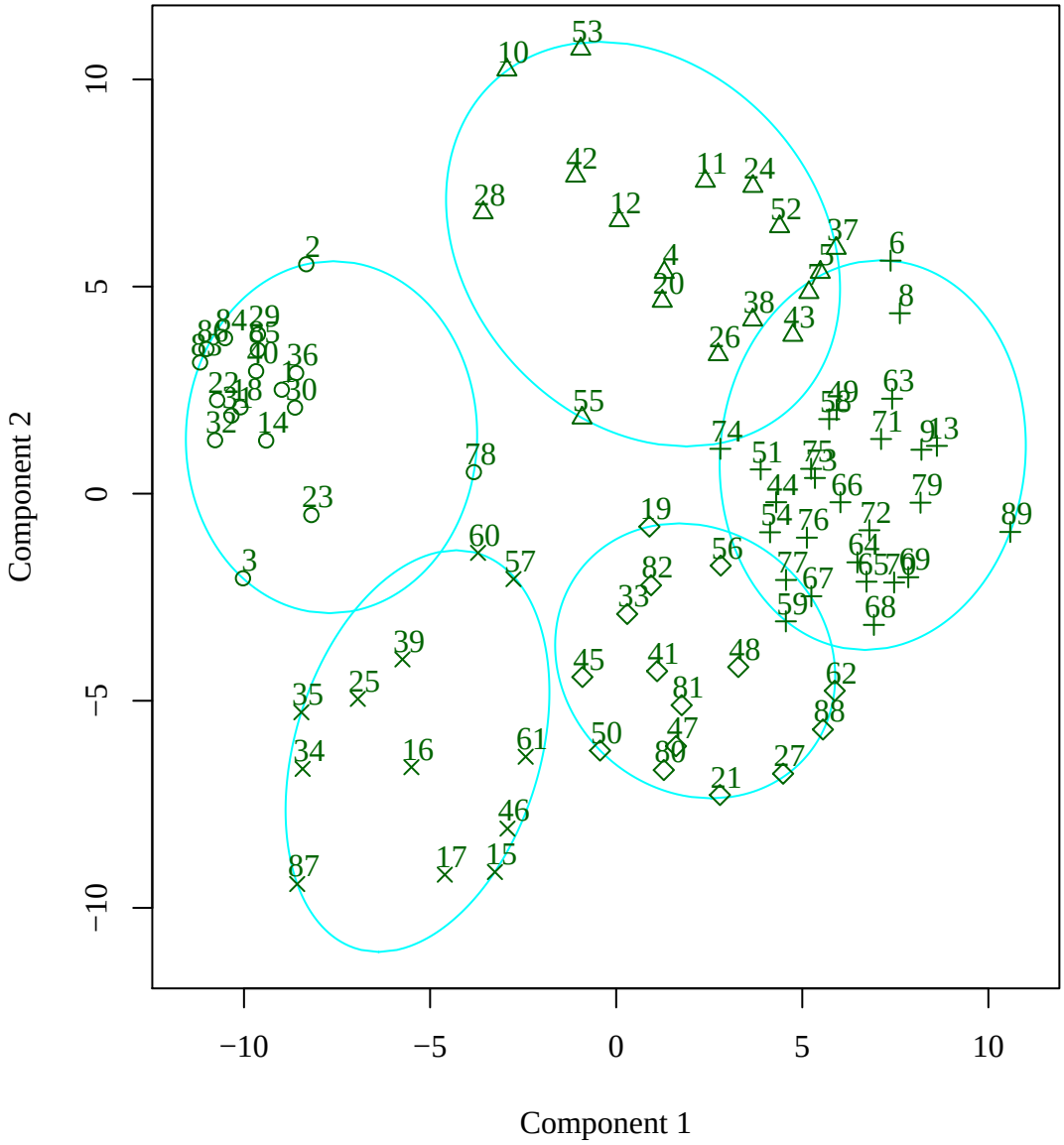
Appendix L

PAM 4-Cluster Solution



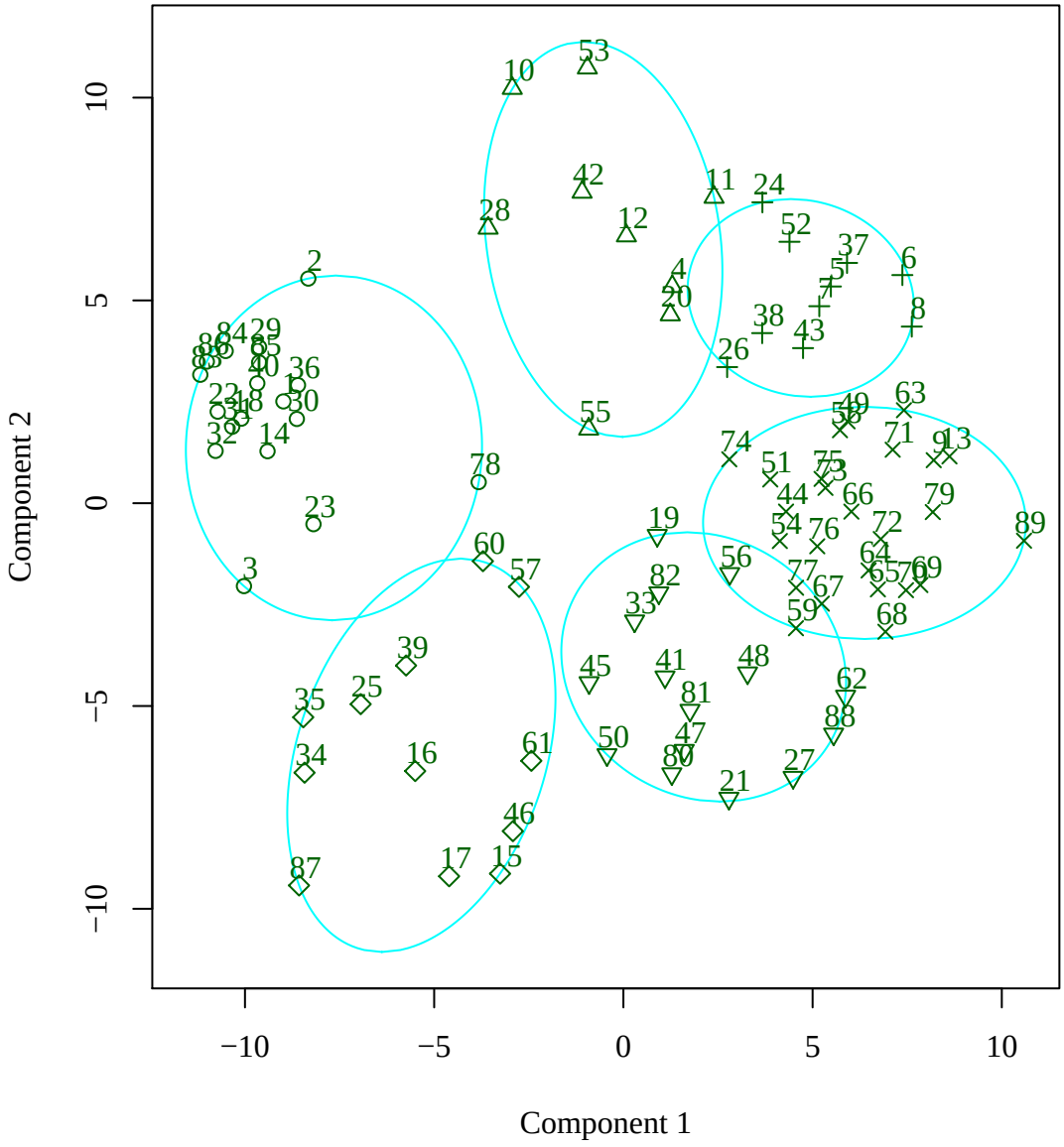
Appendix L

PAM 5-Cluster Solution



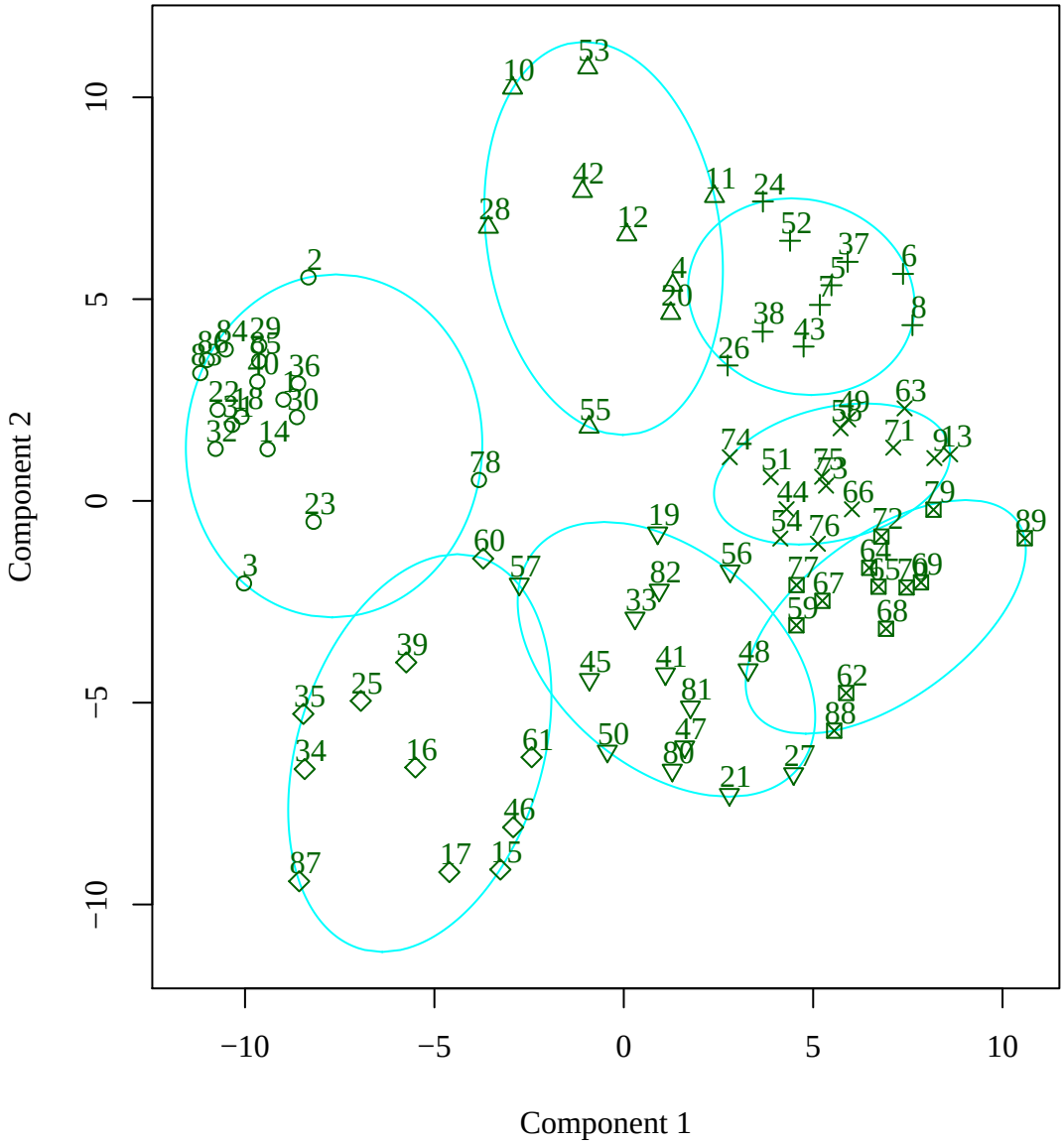
Appendix L

PAM 6-Cluster Solution



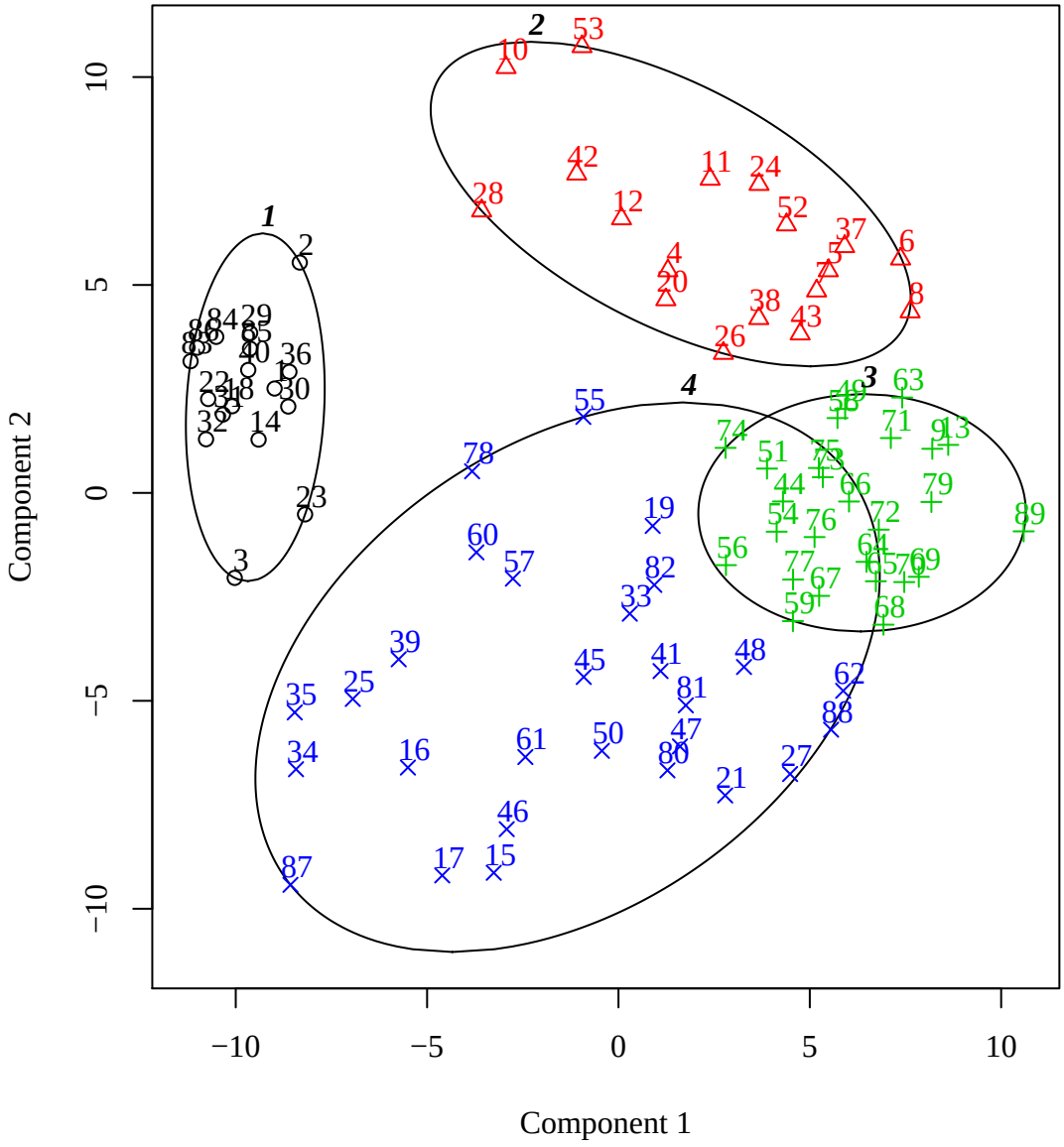
Appendix L

PAM 7-Cluster Solution



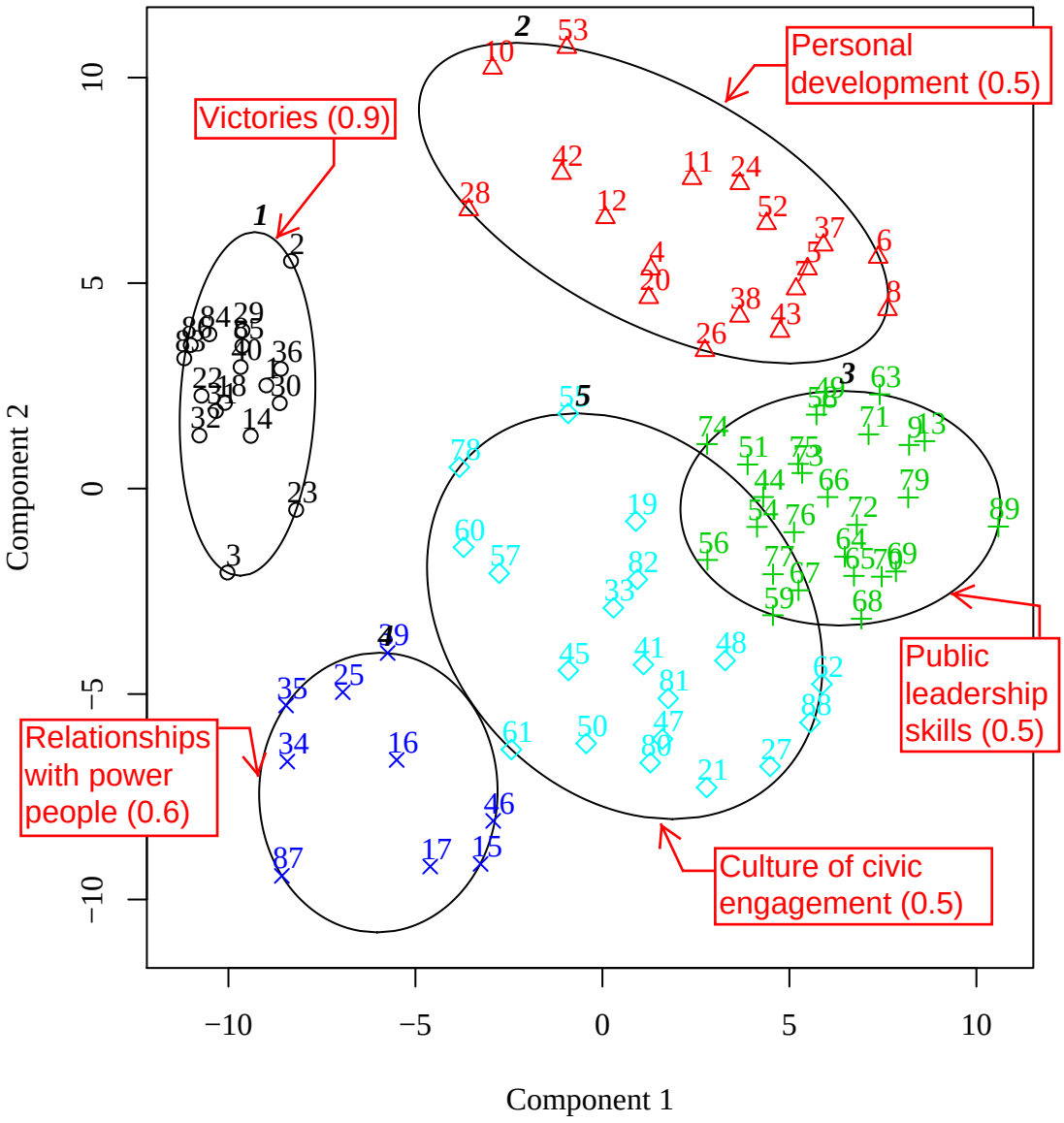
Appendix M

4-Cluster Map for Interpretation Session



Appendix N

Final 5-Cluster Map



Appendix O

Grouped List of Statements – Final 5-Cluster Map

Cluster 1: Victories

1. At Harrington School, MOP helped us get a one way street, a crossing, a bigger parking lot and (soon) a new parking lot. Parents now see there is hope.
2. Healthcare legislation got passed, but it got "ugly" in the process.
3. Healthcare legislation passed because of relationships MOP built; for example, with Senator Bennet and Rep. DeGette.
14. MOP established Parent Liaisons in schools; PLs help connect us to resources in the school.
18. MOP helped get Salud health clinic in Commerce City. It now has its own building and soon 10 more doctors.
22. MOP/PICO helped us win the hospital provider fee bill (100,000 new insured).
23. My child's school principal asks to work more with MOP and with other principals in NNE Denver.
29. Parents won getting classrooms to be like "colleges" at Bruce Randolph School.
30. People got out the vote and defeated the "bad 3" propositions in 2010.
31. Salud clinic provides people jobs.
32. SCHIP was signed by President.
36. The library hours in low-income neighborhoods increased.
40. There is a new transportation network benefiting students in NNE Denver.
83. We were instrumental in getting the first school-based clinic opened in Aurora public schools; it now serves three public & one Catholic school and soon will open a 2nd clinic.
84. We won an island in the road near the school and kids are safer. We had to push the power players to research this issue truthfully.
85. We won housing relocation assistance for mobile home park tenants – people got stipends.
86. Weighted student funding was implemented in DPS.

Cluster 2: Personal development

4. I feel that I do have a voice, I am intelligent and what I have to say is important…even though I don't have a college degree.
5. I get to know people I would otherwise not have known.
6. I learn my own faults & challenges and I learn how to overcome them.
7. I learn to accept that there are people who will always be outside my sphere of influence – and when to move on.
8. I realize that it's not in my self-interest to care just about myself.
10. In a conservative environment, leaders risk being labeled "liberal" if they participate in organizing work. The organizing can alienate some parts of the community.
11. It is now harder for people to "shut me up".

12. Leaders have to "push back" on staff if the staff is moving forward on a different issue that does not meet needs of leaders.
20. MOP opened my eyes to issues that I never paid attention to, for example, immigration & the DREAM act.
24. My kids began asking questions about my involvement in organizing work (this is good).
26. One leader, who never spoke in public, now helps run a meeting. When leaders see that people will listen to them, it changes them.
28. Our church committee pushed the city for a traffic study. We learned to persevere when the solution identified by the city was different than what we hoped for. We compromised.
37. The PICO organizing model doesn't just fix the community, it helps me know what my values are.
38. The skills and involvement we learn with MOP informs other areas in our lives. We are active and engaged community members and better on the job and at church and in school.
42. We are challenged as parents to balance doing organizing work and taking care of our children. Example: going to a MOP meeting.
43. We are personally empowered and become confident leaders.
52. We draw on our faith & values to keep going when we hit barriers in organizing.
53. I find that if both spouses aren't involved in organizing at the same level, there can be tension created by attending so many meetings.

Cluster 3: Public leadership skills

9. If we become board members, we learn new skills and we understand how the organization works.
13. Leaders learn or have learned to have discussions around divisive issues and deal with conflict constructively.
44. We become educated on how to talk, how to get our point across and how to get our story heard in the broader community.
49. We develop analytical and problem-solving skills.
51. We develop skills to put an initiative on the ballot, for example, Aurora recreation center.
54. We find common needs with others and we can identify our self-interest in a common goal.
56. We find that issues are linked. Example: education and healthcare are connected.
58. We gain technical skills (i.e. computer skills).
59. We get informed about community issues.
63. We increase empathy and compassion, even for adversaries.
64. We learn how the political process and system works.
65. We learn how to do research on community problems.
66. We learn how to fight for an issue even if we are not personally impacted…because it's the right thing to do.

67. We learn how to interact with someone who has more power and how to learn from and influence them. Example: decision-makers.
68. We learn how to listen to people's stories, clarify their needs and identify their issues.
69. We learn how to recruit other people to participate.
70. We learn skills like writing press releases and speeches.
71. We learn to compromise (including with MOP staff) regarding an issue cut or strategy.
72. We learn to develop power and identify power, both personally and as a group.
73. We learn why voting matters.
74. We learned about data on how our schools are performing.
75. We learned about the process of constructing a new recreation center. Example: zoning laws.
76. We learned that there are multiple solutions to an issue and that one solution can sometimes address many smaller issues. Example: recreation center.
77. We learned to be strategic and to identify where we have power to impact the healthcare bill. We identified specific issues like affordability that had to be included in the final bill.
79. We meet and learn to work with people who have different viewpoints & beliefs.
89. Within faith communities, MOP's organizing challenges us to examine our morality and spirituality and how it can inform our vote.

Cluster 4: Relationships with power people

15. MOP had a relationship with Senator Bennett because of the work we did while he was superintendent of DPS.
16. MOP has name recognition with members of the school board, city council, state legislature, congress and with their staff members.
17. MOP has relationships with high profile partners and stakeholders, for example, Kaiser, nurses' association & HEAA.
25. Neighbors viewed MOP leaders as a voting information resource, especially on the "bad 3" propositions in 2010 (60, 61 and 101).
34. Some people hesitate to join MOP & only listen to see who you are and if you are serious.
35. The healthcare committee was challenged to "keep going" after SCHIP vetoed twice.
39. The Sun Valley Coalition committee is seen as a resource for power for residents at Decator Place. Residents come to committee members when they need an issue addressed by Decator Place managers.
46. We can empower others to improve their work and report more accurately on the Sun Valley Coalition. Example: professional reporter Tina Griego.
87. When MOP leaders publically speak out on an issue, we sometimes come under attack. Example: Susan testifying before Congress on SCHIP.

Cluster 5: Culture of civic engagement

19. MOP leaders become part of democratic process.
21. MOP puts a "face" on issues and data, humanizes problems in the community.
27. Organizing fosters understanding of who is our community – church, neighborhood, groups near the church.
33. Small groups come together to build more power.
41. We are able to educate others about issues and help to shape their views. Example: classes or forums on healthcare.
45. We build a sense of community within LOCs and with all of MOP and we support each other when there are barriers in the organizing work.
47. We can turn out large numbers of people to demonstrations, actions and press conferences.
48. We come to realize the needs & self-interests of politicians.
50. We develop relationships with the media, including reporters.
55. We find solutions to problems previously thought to be unwinnable.
57. We found out that city council is happy when community members show up and talk with them.
60. We get to know other parents so we can better our children's education and increase communication.
61. We help public officials develop a better sense of responsibility and accountability to the people. Example: city officials listening to parents.
62. We identify groups and people who are in opposition to our agenda.
78. We make systemic change.
80. We participate in non-partisan electoral work including doorknocking and phonebanks.
81. We see that MOP provides childcare and food at meetings which creates a pathway for parents to get involved.
82. We solve problems (even if we don't speak English).
88. When we had difficulties, we followed the PICO principle "All organizing is reorganizing" to keep the momentum going.