The Evaluator as Cartographer:

Technology for Mapping Where We're Going and Where We've Been

William M.K. Trochim
Cornell University

Paper Presented at the 1999 Conference of the Oregon Program Evaluators Network, "Evaluation and Technology: Tools for the 21st Century", Portland, Oregon, October 8, 1999.

DRAFT: Not for quotation, citation or dissemination

Introduction

The renowned organizational psychologist Karl Weick used to tell an anecdote that illustrates the critical importance of maps to a group or organization.[1]

> A group of mountain climbers was in the process of ascending one of the most daunting peaks in the Alps when they were engulfed by a sudden snow squall. All were experienced climbers and each had their own idea of the direction they should go in to get back to the base camp. They wandered around for some time, arguing about which way to go, while their circumstances became more dire and threatening with each moment of indecision. Finally, one of the climbers dug around in their backpack and found a map. Everyone huddled around the map, studied it, and quickly determined their direction. Several hours later, they arrived safely at the camp. While they were warming themselves around the fire, regaling each other with the story of their near misadventure, one of the climbers picked up the map they had used to descend the Alps. On looking at it more carefully, they realized it was actually a map of the Pyrenees!

The map provided the group with an externalized organizing device around which they could reach consensus. It gave them an apparent sense of direction and enabled them to formulate a coherent plan. It led them to take action in a concerted manner. But it was the *wrong* map. Perhaps this group was simply lucky. The way in which their map was wrong happened to provide one possible "correct" path for them. But the story makes it clear that the group was going nowhere anyway. Without some type of device to enable them to coalesce as a group they would almost certainly have perished in the confusion and conflict. Of course, we would prefer that our mountain climbers have the "right" map. But, in a storm, even the wrong map can sometimes lead to better decision making than no map at all.

In an intriguing and stimulating volume several decades ago, Nick Smith (1981) described a number of interesting metaphors for thinking about evaluation work. He showed that it was useful to construe what evaluators did as analogous to what other professions and fields engaged in, sometimes seemingly even very unrelated fields. For instance, he likened aspects of evaluation to what lawyers do when they engage in a trial or what art critics do when they review artwork. He did not mean to suggest that evaluators really were lawyers or art critics, only that some aspects of evaluator's work shared important characteristics with these others.

This paper seeks to extend that notion to a field he did not consider then: cartography. It suggests that one way to think about evaluation is to view it as akin to mapmaking. Like the cartographer, the evaluator gathers information, albeit not geographical in nature[2]. Like the cartographer, evaluators analyze and represent that information, making decisions about how best to draw it, minimize biases and depict perspectives. Like the cartographer, evaluators hope that their representations are useful for guiding others and helping them make more informed decisions. And, in the two decades since Smith's work, we have even seen a form of analysis and representation emerge in evaluation that is called "mapping." These are not geographic maps, but maps of ideas and maps of data. The evaluator facilitates the creation of these maps as an extension of and alternative to more traditional tabular, numerical and textual forms of data representation. And, the evaluator hopes that, like geographic maps, their concept and data maps will help to inform and guide others and enable them to make better decisions.

---

[1] I recall hearing this story but do not have an exact citation. I have done my best to be accurate to the original details. If anyone has a citation for this, please contact me.

[2] The reader should note that this paper does not address in any way the important and growing use of geographic information systems in evaluation research. The "mapping" the metaphor refers to is intended to be the data and concept maps of the type described here.

As with Smith's metaphors, there is no intent here to suggest that evaluation and cartography are the same thing or that one is in any way a subset of the other. There are many ways in which the two fields are distinct. But, this paper intends to show that, especially when it comes to the newer mapping techniques emerging in evaluation, the evaluator will benefit by considering where the cartographers have been and where they appear to be going. Evaluation-cartographers will be better informed and prepared if they learn about the ways cartographers consider issues of context, purpose, methodology, bias, and representation in their own field. This paper hopes to make a beginning in this interdisciplinary endeavor.

Maps and Concept Maps

The idea of a "map" goes back into the farthest reaches of human history, predating even historical records (Wilford, 1982). The impulse to depict visually the distinction between *here* and *there* or to direct one in getting from place to place is a fundamental human trait. The historical record suggests that "…the map evolved independently among many peoples in many separate parts of Earth" (Wilford, 1982). Throughout history maps have played important roles in providing direction, demarcating property, establishing claims, and demonstrating the powers of states.

When most people think of maps, they think of geographical depictions. The naïve assumption is that a map is as accurate a representation of some external physical reality as the cartographer could make it. We will see that this is not the view of most contemporary cartographers, but it is one that persists in the general public.

We can distinguish between several different fundamental types of maps to illustrate both the evolution of thinking about map structure and to show how far we have moved from the idea of the geographical fundamentalism usually association with cartography. To do so, we will develop a simple classification system that will probably not outlive this paper, but that might help orient us to some meaningful distinctions.

Let's begin by making a simple distinction between the "base" of a map and any additional data that is represented upon this base. Furthermore, let's distinguish between whether the base is a geographic one or whether it represents some other relational arrangement. We then have four broad types of maps:

	No Data	Data
Geographical Base	geographic map	geographic data map
Relational Base	relational map	relational data map

Any geographical map -- a highway map, geological survey map, and most maps from a standard atlas -- can be considered geographically "isomorphic". That is, there is a presumed one-to-one relationship between the information depicted on the map and some external physical reality. This isomorphism can never be exact -- it is always an errorful signification. For instance, the star symbol the highway map shown in Figure 1 indicates where my house is located geographically. The star itself is only a sign that represents the house -- I don't live in a star-shaped house, and my house is not nearly as large as depicted on this map -- but the map assumes that there is a one-to-one correspondence (i.e., isomorphism) between the objects on the map and something in geographical reality. If you navigate to the place indicated by the map, you will find my house. The correspondence is "star" = "house" and "line" = "road" and "blue line" = "water" and so on.

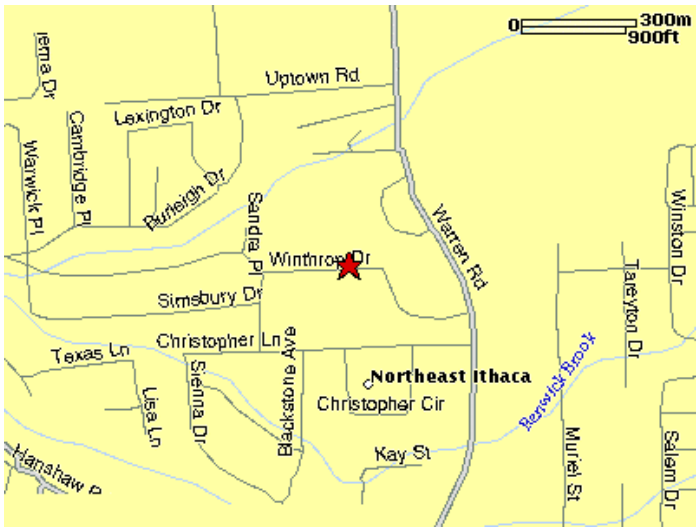


Figure 1. A first-level map with geographic isomorphism.

A second-level map is one that retains a geographical base but depicts data upon it. We might term this type of map a "geographical data map" one. The idea is to represent some characteristic in reference to the geographical base. For instance, consider the map in Figure 2 that depicts earthquakes in the contiguous 48 states. If you were to use this to navigate geographically to a location with white on the map (indicating a high-risk area), you would not be likely to "find" an earthquake in the same sense that you would "find" my house if you follow the geography implicit on Figure 1. In most data maps, the "data" is not in itself geographical. It can simply be represented well with respect to geography.

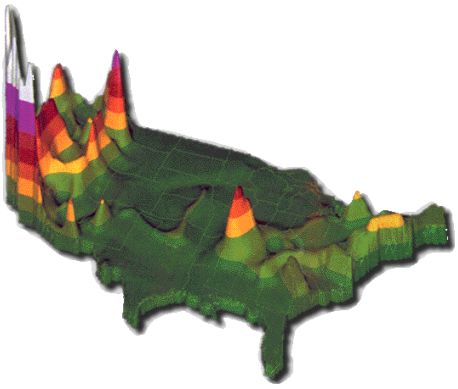


Figure 2. 1989 Computer-generated map showing earthquake-prone areas. High-risk areas appear as white peaks. Courtesy of Melvin L. Prueitt, Los Alamos National Laboratory. Data from the U.S. Geological Survey.

The data can be even more abstract than earthquakes. For instance, we might map specific crimes onto a geographical base as shown in the map for Evansville, Indiana in Figure 3. Again, while there is a geographical base to the data depicted on the map, one would not expect to be able to go to a specific dot on this map and find the indicated event. The geographic base is used as a frame for the depiction of the crime data in a visually meaningful manner.

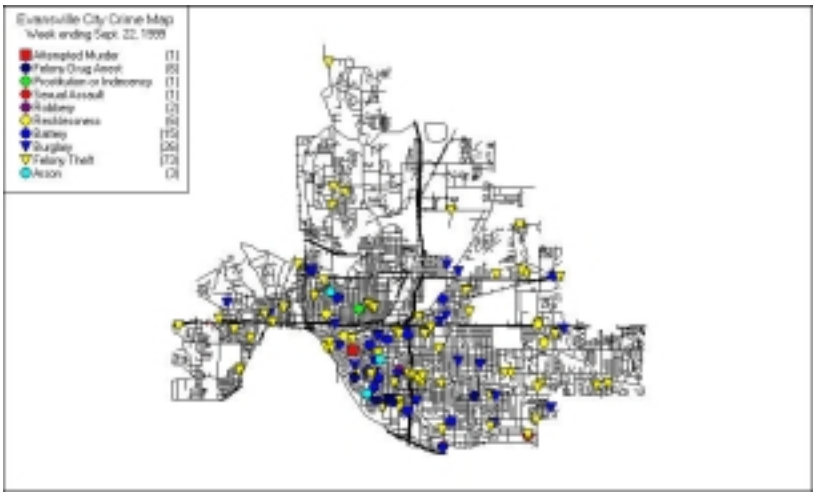


Figure 3. Crime map for Evansville, Indiana, representing crimes for the week ending September 22, 1999.

One more level of abstraction. In Figure 4 we see crime data aggregated by state. Here, the data is even more abstract and highly aggregated than in Figure 3. We would not expect that by going to the state of Florida we would "see" in any direct manner the higher crime reflected in the dark fill for that state. The geographic base is used as a framework for representing information on another variable or measure. It is important to recognize that the data being represented on this map represents considerable judgment. What is meant by a "serious" crime for purposes of the map? How accurately can we assume such crimes are actually measured? How are the levels of crime divided into the five categories depicted in the different shades? How might the map look different if the data was shown by county?

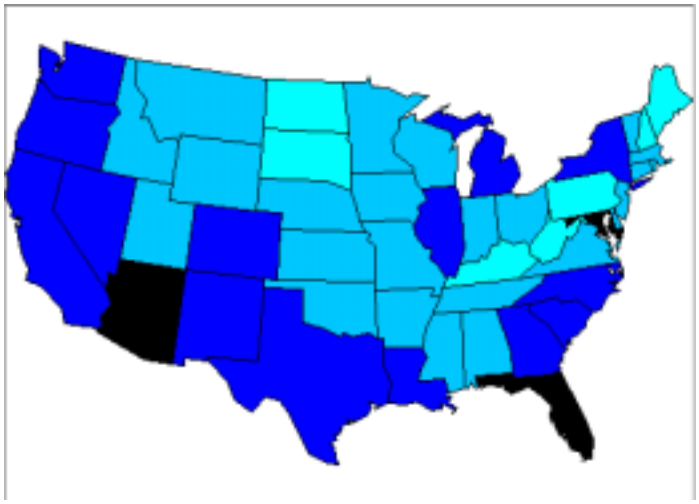


Figure 4. Serious crime by state (darker areas indicate higher serious crime rates.

The type of geographical data map shown in Figure 4 is an invention of the past few centuries. The idea of using geographical maps as a basis for depicting other non-geographic information is a distinctly "modern" one. Note that we can move to increasingly abstract characteristics as the objects of representation. For instance,

instead of measuring crime, we could measure individuals' perceptions of safety. We could conduct a public opinion survey asking people to rate subjectively the degree to which they feel safe in their home communities and graph that information onto a geographic base. In this case, the respondent's perception of safety may bear little or no resemblance to the pattern of reported crime for the same geographical area.

If we move away from the geographical base, we come to the third type of map which will be termed here a "relational map." Here, the base is not a geographical one. For instance, if we asked a person to "draw" a map of different types of crimes, placing crimes that they think are more similar to each other closer together on the map, we would have a relational representation. The "base" for this map would consist not of geographical entities but, in this case, of crimes. These crimes would be located in proximity to one another based on a person's *judgment* of their similarity. The implicit "landscape" for such a map is the person's subjective perceptual relational arrangement of crimes and crime types.

Any map that uses non-geographic relationship information as the base can be classified as a relational map. If the intent of the map is to represent accurately some implicit cognitive reality, we can refer to it as a *relationally isomorphic* map. Here, each symbol has a one-to-one relationship to an idea or construct, and the arrangement of symbols on the map shows how the ideas are thought to be interrelated.

In the past several decades we have seen a large number of these types of maps evolve. They range from the "mind maps" recommended by Tony Buzan (Buzan, 1993) and illustrated in Figure 5 to the concept maps developed by Joe Novak (Novak and Gowin, 1985, Novak, 1993) and exemplified in Figure 6. In both types of maps, the intent is to represent a person's thinking pictorially. The relational structure is based on lines that depict relationships between ideas. No attempt is made to develop a meaningful Euclidean framework on which to depict relationships -- it is the connecting lines that carry the relational information.

In most cases, these types of maps are constructions of individuals. Where groups are involved, there is no algorithm for aggregating -- the group must do that interactively.

Figure 5. A stylized mind map in the style of Tony Buzan (Buzan, 1993) showing the linkages among researchers' ideas for a research project.

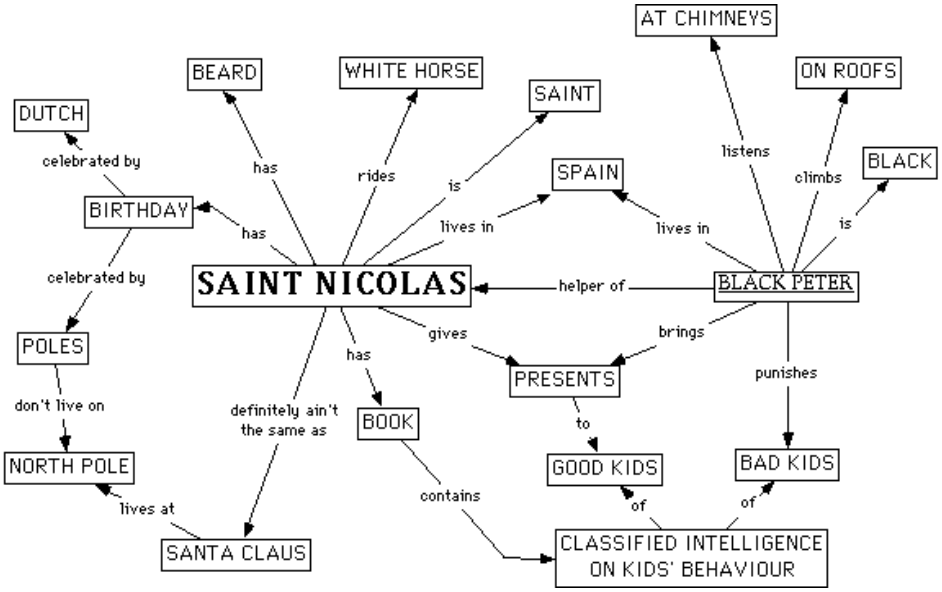


Figure 6 A relationally isomorphic concept map in the style of (Novak and Gowin, 1985) on the topic of St. Nicolas..

We come finally to the types of maps described in this paper as concept maps. Illustrations of these types of maps are provided later. These are also relational maps. But in these maps the relational base provides the structure for carrying or representing additional data. In this sense, we can describe these as *relational data maps*. These concept maps are usually constructions of groups. They utilize

mathematical algorithms that aggregate individuals' judgments of the similarity among ideas and represent the ideas in symbols arrayed in Euclidean space. In this sense, these maps are like geographical maps in that the distance between symbols is meaningfully interpreted as an empirical estimate of the semantic distance between ideas.

The point of this classification system, other than providing us with useful terms to distinguish different types of maps, is to show that there is a continuum between the traditional geographic maps and the more recently evolved conceptual ones. They are less distinct than might at first appear to be the case. Given this relationship between traditional geographical maps and the concept maps that are the subject of this paper, we might look to the cartographic profession and their current methodological discussions for insights that help us understand better the issues of quality, interpretability and validity involved in concept mapping.

Before doing so, the basic procedure for producing a concept map is described, and a case study example presented in some detail. With this prologue, we can examine some of the issues or relevance that contemporary cartography is grappling with.

Concept Mapping

Concept Mapping

Concept mapping is a process that can be used to help a group describe its ideas on any topic of interest (Trochim, 1989a) and represent these ideas visually in the form of a map. The process typically requires the participants to brainstorm a large set of statements relevant to the topic of interest, individually sort these statements into piles of similar ones and rate each statement on some scale, and interpret the maps that result from the data analyses. The analyses typically include a two-dimensional multidimensional scaling (MDS) of the unstructured sort data, a hierarchical cluster analysis of the MDS coordinates, and the computation of average ratings for each statement and cluster of statements. The maps that result show the individual statements in two-dimensional (x,y) space with more similar statements located nearer each other, and show how the statements are grouped into clusters that partition the space on the map. Participants are led through a structured interpretation session designed to help them understand the maps and label them in a substantively meaningful way.

The concept mapping process discussed here was first described by Trochim and Linton (1986). Trochim (1989a) delineates the process in detail and Trochim (1989b) presents a wide range of example projects. Concept mapping has received considerable use and has been used to address substantive issues in social services (Galvin, 1989; Mannes, 1989), mental health (Cook, 1992; Kane, 1992; Lassegard, 1993; Marquart, 1988; Marquart, 1992; Marquart et al, 1993; Penney, 1992; Ryan and Pursley, 1992; Shern, 1992; Trochim, 1989a; Trochim and Cook, 1992; Trochim et al, in press; Valentine, 1992), health care (Valentine, 1989), education (Grayson, 1993; Kohler, 1992; Kohler, 1993), educational administration (Gurowitz et al, 1988), training development (McLinden and Trochim, in press) and theory development (Linton, 1989, Witkin and Trochim, 1996). Considerable methodological work on the concept mapping process and its potential utility has also been accomplished (Bragg and Grayson, 1993; Caracelli, 1989; Cooksy, 1989; Davis, 1989; Dumont, 1989; Grayson, 1992; Keith, 1989; Lassegard, 1992; Marquart, 1989; Mead and Bowers, 1992; Mercer, 1992; SenGupta, 1993; Trochim, 1985 , 1989c, 1990, 1993).

How a Concept Map is Produced

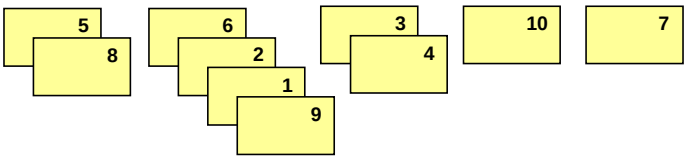
Concept mapping combines a group process (brainstorming, unstructured sorting and rating of the brainstormed items) with several multivariate statistical analyses (multidimensional scaling and hierarchical cluster analysis) and concludes with a group interpretation of the conceptual maps that result.

In the typical situation, concept mapping begins with the formulation of a focus statement that guides and delimits the scope of the map. A set of statements that address this focus statement is then produced, usually through some form of brainstorming[3]. Two types of data are typically collected with respect to these statements. First, each participant is asked to sort the statements into piles of similar ones, an unstructured similarity sort. The sort is required -- a concept map cannot be produced without sorting data. Second, each participant is usually (although this is not a requirement) asked to rate each statement on one or more variables. Most typically, each statement is rated for its relative importance, usually on a 1 (Relatively Unimportant) to 5 (Extremely Important) scale. The rating information is not used to produce the base of the map itself -- it is only used as an overlay on a map that was constructed from sort data.

The first step in the analysis involves transforming each participant's sort into quantitative information. The challenge here is to find a way to "aggregate" or combine information across participants given that different individuals will have different numbers of sort groups or piles. The solution is to place each sort result into the same-sized matrix. Figure 7 illustrates this for the simple example of a single participant and a ten-statement sort. This person sorted the 10 statements into 5 piles or groups. Other participants may have had more or fewer groups. but all sorted the same *number* of statements, in this example, ten. So, we construct a 10x10 matrix or table of numbers. For each individual, the table is a binary one consisting only of 0s and 1s. If two statements were placed together in a pile, their corresponding row and column numbers would have a 1. If they weren't placed together, their joint row-column value would be a 0. Because a statement is always sorted into the same pile as itself, the diagonal of the matrix always consists of 1s. The matrix in symmetric because, for example, if statement 5 is sorted with statement 8 then it will always be the case that statement 8 is sorted with 5. Thus, the concept mapping analysis begins with construction from the sort information of an NxN binary (where N = the number of statements), symmetric matrix of similarities, X_{ij}. For any two items i and j, a 1 is placed in X_{ij} if the two items were placed in the same pile by the participant, otherwise a 0 is entered (Weller and Romney, 1988, p. 22).

[3] The concept mapping methodology doesn't know or care how the statements are generated. They could be abstracted from existing documents, generated by an individual, developed from an interview transcript, and so on. All that the method requires is that there is a set of statements. Of course, one's interpretation of the maps would depend critically on how the statements were generated. The current version of the Concept System software allows for up to 200 statements in a map, although it would be rare that a participant group would be comfortable dealing with any more than 100 or so.

Sort Cards for 10 statements



Binary Square Similarity Matrix for one Person

	1	2	3	4	5	6	7	8	9	10
1	1	1	0	0	0	1	0	0	1	0
2	1	1	0	0	0	1	0	0	1	0
3	0	0	1	1	0	0	0	0	0	0
4	0	0	1	1	0	0	0	0	0	0
5	0	0	0	0	1	0	0	1	0	0
6	1	1	0	0	0	1	0	0	1	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	1	0	0	1	0	0
9	1	1	0	0	0	1	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

Figure 7. Transforming sort data into a binary square similarity matrix.

With this simple transformation of the sort into matrix form, we now have a common data structure that is the same size for all participants. This enables us to aggregate across participants. Figure 8 shows how this might look when aggregating sort results from five participants who each sorted a ten-statement set. In effect, the individual matrices are "stacked" on top of each other and added. Thus, any cell in this matrix could take integer values between 0 and 5 (i.e., the number of people who sorted the statements); the value indicates the *number* of people who placed the i,j pair in the same pile. Thus, in this second stage, the total NxN similarity matrix, T_{ij} was obtained by summing across the individual X_{ij} matrices.

Total Square Similarity Matrix Across Five Participants

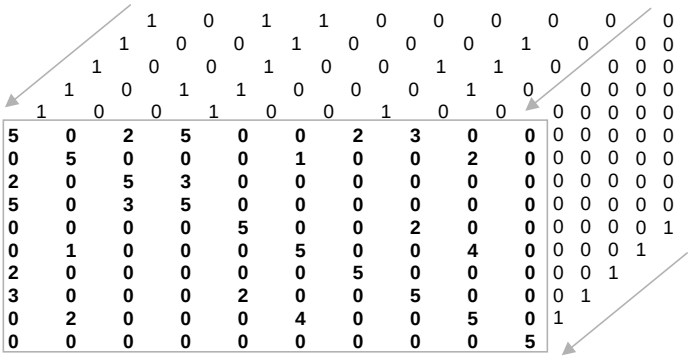


Figure 8. Aggregating sort data across participants.

It is this total similarity matrix T_{ij} that is analyzed using nonmetric multidimensional scaling (MDS) analysis with a two-dimensional solution. The solution is limited to two dimensions because, as Kruskal and Wish (1978) point out:

> *Since it is generally easier to work with two-dimensional configurations than with those involving more dimensions, ease of use considerations are also important for decisions about dimensionality. For example, when an MDS configuration is desired primarily as the foundation on which to display clustering results, then a two-dimensional configuration is far more useful than one involving three or more dimensions (p. 58).*

The analysis yields a two-dimensional (x,y) configuration of the set of statements based on the criterion that statements piled together by more people are located closer to each other in two-dimensional space while those piled together less frequently are further apart. The similarity matrix input and the most basic "point map" output are shown in Figure 9.

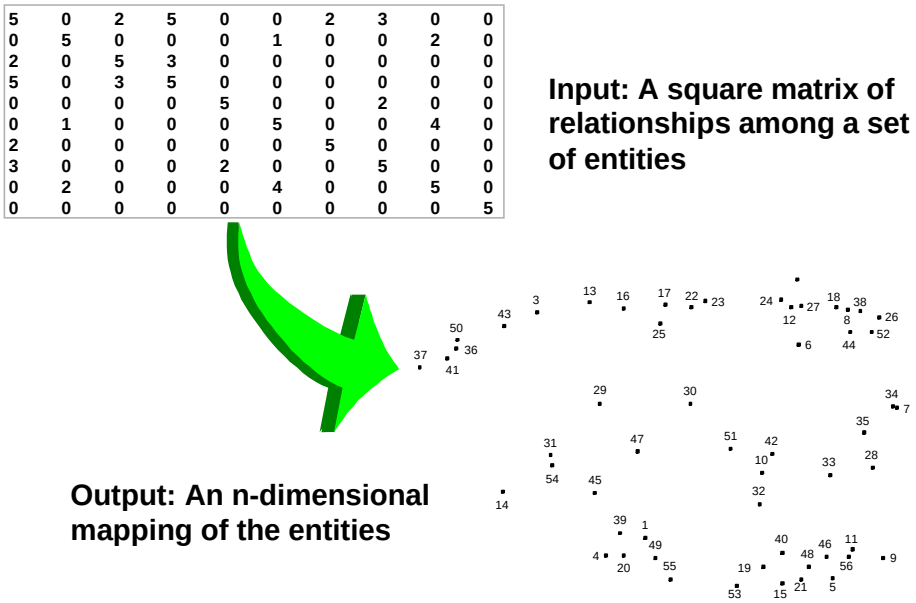


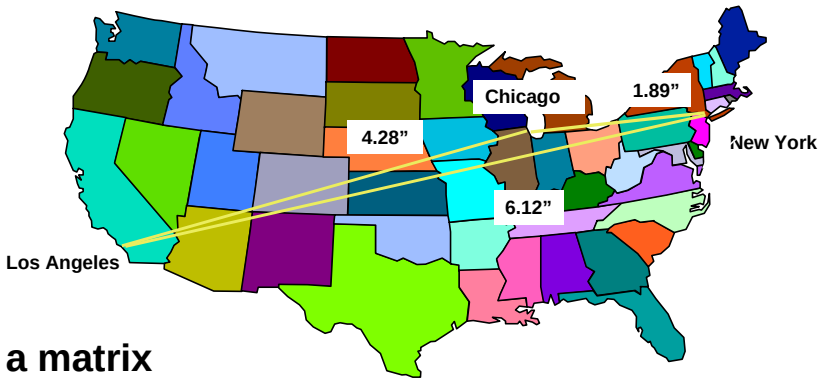
Figure 9. Input and output of the map analysis.

Multidimensional scaling (MDS) is the analytic procedure that accomplished the basic mapping shown in Figure 9. How does it do this? There are numerous mathematical descriptions of the process (Davison, 1983, Kruskal and Wish, 1978) that will not be repeated here. Instead, we attempt to provide a nonmathematical explanation that will hopefully provide some insight regarding how MDS accomplishes its work.

One way to accomplish this intuitively is to think about the *opposite* of what MDS accomplishes. As described above in Figure 9, MDS takes a square matrix of (dis)similarities[4] for a set of items/objects as input and produces a map[5] as output. To see how it does this, think about the much more intuitive task of going in the opposite direction -- starting with a map and, from it, producing a table of (dis)similarities. Figure 10 shows a map of the United States with three major cities indicated. The cities are the "objects", the points on this map that are analogous to the statements on a concept map. How would you produce, from a two-dimensional map like this, a table of (dis)similarities? The simplest way would be to use a ruler to measure the distances (in whatever units) between all possible pairings of the three points. The figure shows the distances obtained (in inches) and the table of dissimilarities. This is the opposite of what MDS does, but it is a common task and easily understood by anyone who has ever worked with a map.

[4] The term (dis)similarity is used in the MDS literature to indicate that the data can consist of either dissimilarities or similarities. In concept mapping, the data is always the square symmetric similarity matrix that is generated from the sorting data.

[5] The "map" is the distribution of points that represent the location of objects in N-dimensional space. in concept mapping, the objects are the brainstormed (or otherwise generated) statements and the map that MDS produces is the point map in two dimensions.



To get a matrix of dissimilarities from a map

	NY	CH	LA
NY	--	1.89”	6.12”
CH	1.89”	--	4.28”
LA	6.12”	4.28”	--

Figure 10. Doing the "opposite" of MDS -- moving from a map to a table of (dis)similarities.

Now, consider how MDS works in the context of this example. MDS would start with a matrix of distances between the three cities and would produce a map that shows the three cities as points. Figure 11 shows how this would work. We begin with a common table, easily obtained from an almanac or atlas. The table shows the straight airline distances (in miles) between the three cities. The goal of MDS would be to convert this information into a map. We'll limit the result to two-dimensions because that is the solution we typically use in concept mapping. How might you do this manually? The airline distances in the table range from 713 to 2451 and those units are most likely inconvenient for drawing a map on a piece of paper. As a first step, you might convert the airline distances into a unit of measurement that is more convenient for graphing. In the figure, we convert to inches using a scale of 1 inch for every 500 airline miles. Now that we are operating in inches, we can more easily work on a sheet of paper. The object is to place three points on the piece of paper so that the distances between the points best represents the distances in inches in the table. You would start by placing one point. In this case, let's assume you place a point on the paper to represent Los Angeles. Next, you place the second point. It doesn't matter whether you choose New York or Chicago, so let's assume you arbitrarily select New York. Where do you place it? According to the table of distances, you need to place New York 4.90" from Los Angeles (note that it doesn't matter in which direction you go, only that New York is exactly that distance from Los Angeles). The figure shows a ruler that indicates where you would place the NY point.

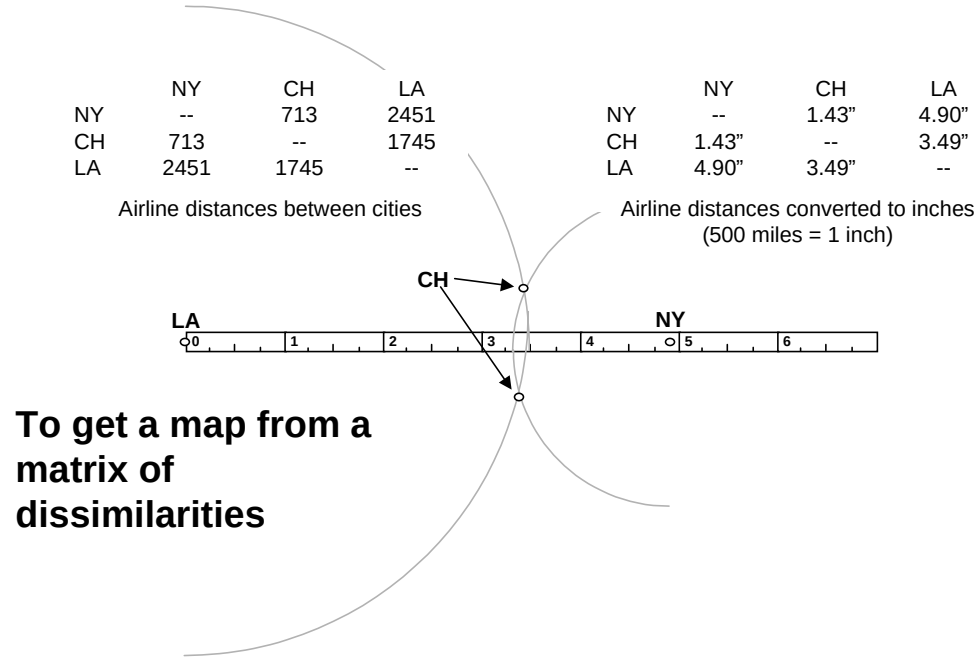


Figure 11. How to manually obtain a two-dimensional map from a matrix of dissimilarities.

All that remains is to place the point for the third city, in this case, Chicago. Where should it go? It must simultaneously meet the condition of being both 3.49" from Los Angeles *and* 1.43" from New York. Figure 11 shows how you might meet these conditions using a compass to draw a semi-circle from Los Angeles at 3.49" and one from New York at 1.43". The semi-circles interact at two points and either of these would be equally good locations for the Chicago point if the object is to represent the distances.

With only three cities, it's a pretty simple matter to construct a map from a table of distances. But what if you had four, or ten, or even 100? The process would rapidly become tedious and, after a few points would become well nigh impossible to accomplish manually. In concept mapping we usually have lots of ideas, sometimes as many as 100 or even 150 that need to be mapped. The input for the mapping -- the analogy to the airline distance table -- is the matrix of similarities among statements that is obtained from the sort task as described in Figure 8. The output is the point map of the statements. MDS accomplishes mathematically a process analogous to what you would do manually in our simple three-city example, except that it could do so for a table that includes a hundred cities or more.

There are several important insights from this simple description of what MDS does. MDS does not know compass directions. It does not know North from South. In the example in Figure 11, MDS would just as happily place Chicago in either of the two locations. It might just as easily place Los Angeles on the right as on the left. This means that when you look at a concept map generated by MDS, direction on the map is entirely arbitrary. You could take a concept map and flip it horizontally or vertically and/or rotate it clockwise or counterclockwise any amount *and this would have no effect on the distances among the points*. The simple exercise shows that MDS yields a relational picture and is indifferent to directional orientation.

In our three-city example, there will always be a two-dimensional solution that will represent the table exactly, with no error. When we move to larger numbers of points this will no longer be the case -- we will not be likely to be able to represent the (dis)similarities exactly. Some (dis)similarity matrices will be able to be represented

more exactly in two dimensions than others will. In MDS, we estimate the overall degree of correspondence between the input (i.e., the (dis)similarity matrix) and the output (i.e., distances between points on the map) using a value called the Stress Value. A *lower* stress value indicates a *better* fit, higher stress means the fit is less exact. In general, you want lower stress, although it is not always clear whether slight differences in stress can be translated into meaningful differences in the interpretability of a map. The normative range for judging stress values in a particular study should be determined from comparisons with similar types of data collected under similar circumstances. We would never expect to get a stress value in a concept mapping study of 100 ideas/statements that is anywhere near as low as what we would obtain from a map of distances among 100 cities! We would also not expect to get as low a stress value if we mapped cities not in terms of their airline distances, but rather in terms of measures that have more variability like crime rates, annual rainfall, or even perceived quality of life. In a study of the reliability of concept mapping, Trochim (1993) reported that the average Stress Value across 33 concept map projects was .285 with a range from .155 to .352. While the stress value has some use in an interpretation, giving you some idea of how accurately your map represents the input data relative to other maps, it is not clear that maps with lower stress are more interpretable or useful than ones with considerably higher stress.

The discussion to this point shows how the concept mapping analysis uses sorting results and MDS to produce the basic "point map" that is the foundation for all other maps. While this is useful in itself, it is helpful to be able to view a concept map at different levels of detail. Just as in geographic mapping, there are times when you want considerable detail (e.g., when hiking or mountain climbing) and other times where grosser-level maps (e.g., when driving on interstate highways) are more useful. The point map generated by MDS is a fairly detailed map. To get to higher-level maps that summarize across some of that detail a procedure known as hierarchical cluster analysis is used. The input to the cluster analysis is the point map, specifically the x,y values for all of the points on the MDS map. Using the MDS configuration as input to the cluster analysis forces the cluster analysis to *partition* the MDS configuration into non-overlapping clusters in two-dimensional space. Unfortunately, mathematicians do not agree on what constitutes a cluster mathematically and, consequently, there is a wide variety of algorithms for conducting cluster analysis, each of them yielding different results. In concept mapping we usually conduct hierarchical cluster analysis utilizing Ward's algorithm (Everitt, 1980) as the basis for defining a cluster. Ward's algorithm has the advantage of being especially appropriate with the type of distance data that comes from the MDS analysis. The hierarchical cluster analysis takes the point map and constructs a "tree" that at one point has all points together (in the trunk of the tree) and at another has all points as their own end points of the "branches". All hierarchical cluster analysis approaches are divided into two broad types, agglomerative and divisive. In agglomerative, the procedure starts with each point as its own branch end-point and decides which two points to merge first. In each successive clustering iteration, the algorithm uses a mathematical rule to determine which two points and/or clusters to combine next. Thus, the procedure agglomerates the points together until they are all in one cluster. Divisive hierarchical cluster analysis works in the opposite manner, beginning with all points together and deciding based on a mathematical rule how to divide them into groups until each point is its own groups. Ward's method is an agglomerative one.

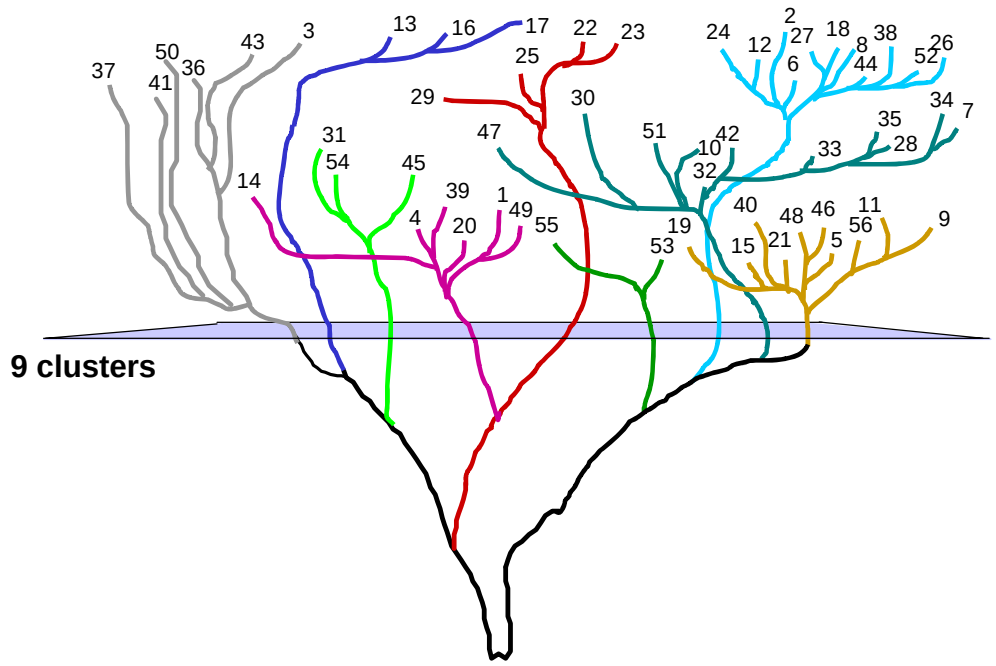


Figure 12. A suggestive hierarchical cluster tree for a point map, showing how a nine-cluster solution would be obtained as a "slice" from the tree.

The kind of thing that is going on in hierarchical cluster analysis is suggested in Figure 12[6]. The numbers show the locations of statements on a point map that was generated by MDS. Each statement is the end-point of a branch. The tree shows how statements get agglomerated and eventually combined onto a single trunk -- a one-cluster solution. By taking vertical slices at different heights of the tree, one can look at different numbers of clusters. The figure shows a vertical slice that would result in nine separate clusters of statements. The resulting nine-cluster solution is shown in two-dimensional concept map form as a "point-cluster" map in Figure 13.

[6] The figure is labeled "suggestive" because it is only meant to convey the idea of a cluster tree visually. There is some deliberate distortion of dimensionality here for graphic purposes.

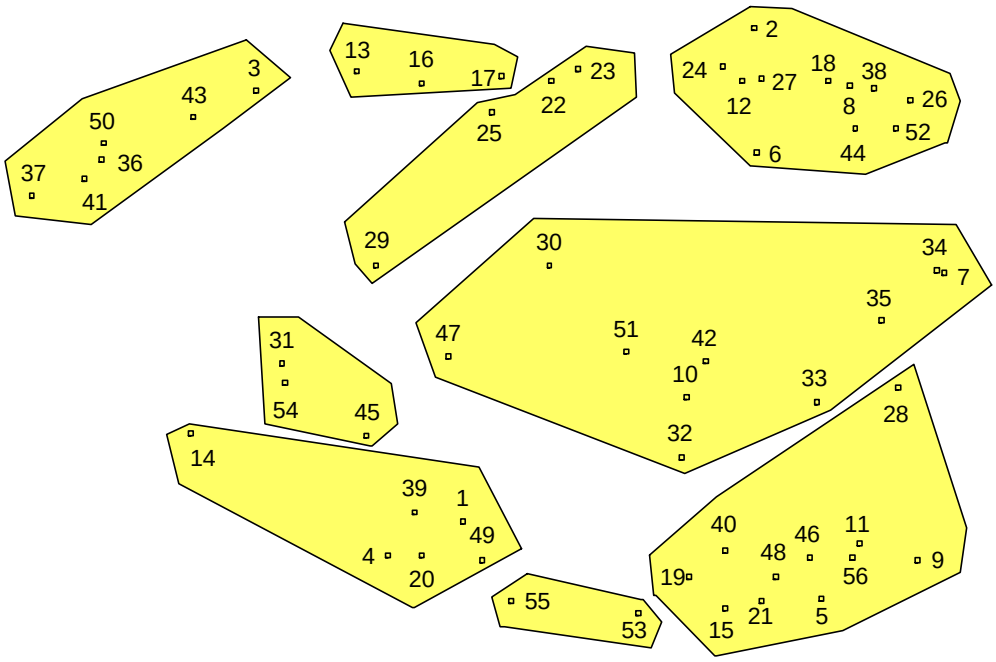


Figure 13. Two-dimensional nine-cluster point-cluster map.

In most concept mapping application contexts it is not useful in interpreting results to show the entire cluster analysis tree. Just as in geographic mapping, it is not feasible to show all levels of detail simultaneously. The geographic cartographer makes decisions about scale and detail depending on the intended uses of the map. Similarly, in concept mapping, the facilitator, often in collaboration with a small group of participants, decides on the number of clusters to use in maps that are reported/presented. There is no simple mathematical criterion by which a final number of clusters can be selected. The procedure typically followed in concept mapping is to examine an initial cluster solution that was the maximum thought desirable for interpretation in this context. Then, successively lower cluster solutions (i.e., successive slices moving down the tree) are examined, with a judgment made at each level about whether the merger seemed substantively reasonable, defensible or desirable. The pattern of judgments of the suitability of different cluster solutions is examined and results in a decision of a specific cluster solution that seems appropriate for the intended purpose of the project. In some projects several such cluster maps are produced to illustrate different levels of aggregation.

All that remains in the core of the concept mapping analysis is the incorporation of any rating data or other measures that may have been gathered. Note that the only data needed to produce the base point and cluster maps described above is the sorting data. Rating data is always used in concept mapping to provide a third-dimension, a vertical overlay that graphically depicts "height" of various areas of the map. For instance, if as part of the data organizing phase, a simple rating of the relative importance of each statement was obtained from participants, one could graph the average importance rating for each statement by extruding[7] a third dimension on each point of the point map. Similarly, the importance of each cluster could be indicated using "layers" for the clusters to indicate the average importance

[7] After years of playing around with this stuff, it seems to be a consensus in concept mapping that a pseudo-three-dimensional rendering is more interpretable than the technically more accurate true three-dimensional representation.

rating of all statements in the cluster. Examples of variations of these maps will provided below.

A Concept Mapping Example

This case study illustrates the use of concept mapping as the foundation for strategic planning for the Oregon Program Evaluators Network (OPEN). It was undertaken as a an informal demonstration project and done in advance of the annual OPEN Conference where the results were to be presented and discussed. OPEN is an organization for Oregonians and Washingtonians involved with, or interested in, program evaluation. It allows members to exchange ideas and information which promotes and encourages high-quality evaluation practices. The founding members of OPEN represent government agencies, universities, and private consulting firms. OPEN's mission is:

> *To provide a regional, interdisciplinary forum for professional development, networking, and exchange of practical, methodological, and theoretical knowledge in the field of evaluation.*

Participants

There were approximately 325 OPEN members on an electronic mailing list who were contacted by e-mail and asked to participate in the study. Since all participation up to the interpretation of maps was done over the World Wide Web, participants self-selected to go to the website and take part. Because brainstorming is anonymous, there is no way of knowing exactly how many participants generated the 83 brainstormed statements. The organizing phase of the process requires that participants logon and, consequently, it is possible to determine sample size exactly. There were 23 people who successfully logged onto the web site, and 22 of these answered at least some of the demographics questions (which only take a few seconds to complete). The sorting and rating tasks are more demanding. Sorting can take as long as 45 minutes to an hour and rating typically requires 10-15 minutes for a set of 83 statements. Of the 23 who logged on, 17 successfully completed the sort and 18 completed the rating task.

Procedure

The general procedure for concept mapping is described in detail in Trochim (1989a). Examples of results of numerous concept mapping projects are given in Trochim (1989b). The process implemented here was accomplished over the World Wide Web using the Concept System© Global web program. This program can be used from any terminal that can access the World Wide Web with any reasonably contemporary web browser such as Internet Explorer or Netscape Navigator (versions 2 or higher). No software needs to be downloaded to the client machine and no applets or other programs are automatically downloaded.

The data collection for the project took place in two phases between September 4, 1999 and October 1, 1999. In the first phase (September 4 - 17), participants were asked to brainstorm statements to a specific prompt. In the second phase (September 20 - October 1), participants were asked to sort and rate the statements and provide basic demographics data.

Phase I: Generation of Conceptual Domain. In the first phase, participants generated statements using a structured brainstorming process (Osborn, 1948) guided by a specific focus prompt that limits the types of statements that are acceptable. The focus statement or criterion for generating statements was

operationalized in the form of the complete-the-sentence focus prompt to the participants:

> *One specific thing I think the Oregon Program Evaluators Network should do over the next five years is…*

The brainstorming interface is illustrated in Figure 14. Participants were only required to point their browsers to the project web page -- no software was required on the participant's machine other than a traditional web browser and web access.

The participants brainstormed 83 statements. The complete set of statements is shown in Table 1. When the brainstorming period ended, two of the principles from the client group were asked to edit the brainstormed statements to assure that spelling and grammar was correct. The website was then set up for Phase II of the project.

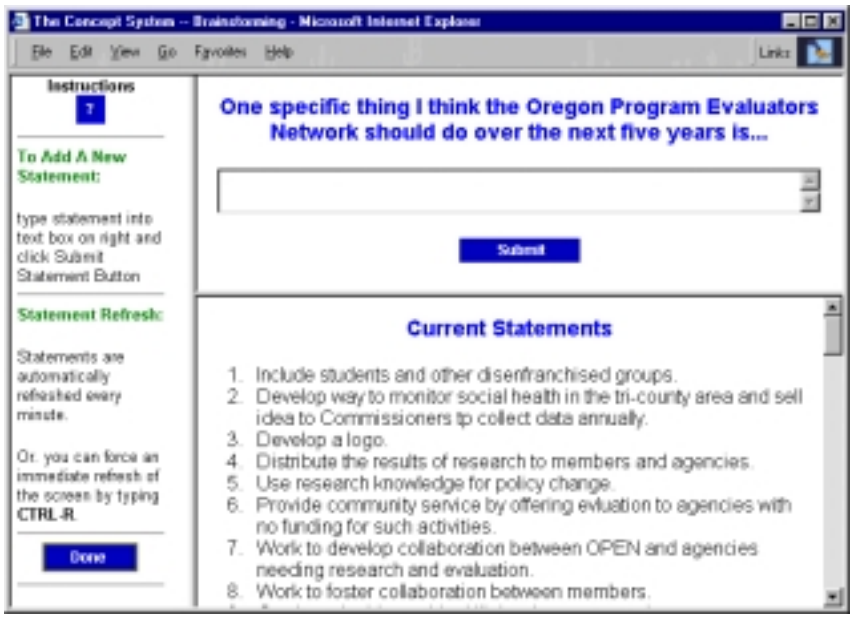


Figure 14. Web-based brainstorming interface for statement generation using the Concept System Global software.

Phase II: Organizing the Brainstormed Statements. As in Phase I, this phase was accomplished entirely over the web. The organizing phase involved three distinct tasks, the sorting and rating of the brainstormed statements, and collection of basic demographic variables. For the sorting (Rosenberg and Kim, 1975; Weller and Romney, 1988), each participant grouped the statements into piles of similar ones. They were instructed to group statements together if they seemed similar in meaning. Each group or pile of statements was given a name by the participant. The only restrictions in this sorting task were that there could not be: (a) N piles (every pile having one item each); (b) one pile consisting of all items; or (c) a "miscellaneous" pile (any item thought to be unique was to be put in its own separate pile). Weller and Romney (1988) point out why unstructured sorting (in their terms, the pile sort method) is appropriate in this context:

> *The outstanding strength of the pile sort task is the fact that it can accommodate a large number of items. We know of no other data collection method that will allow the collection of judged similarity data among over 100 items. This makes it the method of choice when large numbers are necessary. Other methods that might be used to collect similarity data, such as triads and paired comparison ratings, become impractical with a large number of items (p. 25).*

For the rating task, each participant was asked to rate each statement on a 5-point Likert-type response scale in terms of how important the statement is with respect to the future of OPEN. The specific rating instruction was:

> Please rate each statement for how important you think it is for the future of OPEN, where
>
> 1=Relatively unimportant
> 2=Somewhat important
> 3=Moderately important
> 4=Very important
> 5=Extremely important

The demographics that were collected are shown in Table 2. Seventeen participants had complete sorting data and eighteen had complete rating data.

Results

The first and most elemental map that is produced in the analysis is the point map that shows the brainstormed statements in a two-dimensional space. Each point on the map represents a single brainstormed idea. Points that are closer to each other physically are more similar to each other cognitively, that is, they tended to be placed together in the same sort piles by more of the participants. Hierarchical cluster analysis is used to "partition" the points into graphically adjacent and proximal groups or clusters. The point map overlaid on the eight-cluster solution is shown in Figure 15.

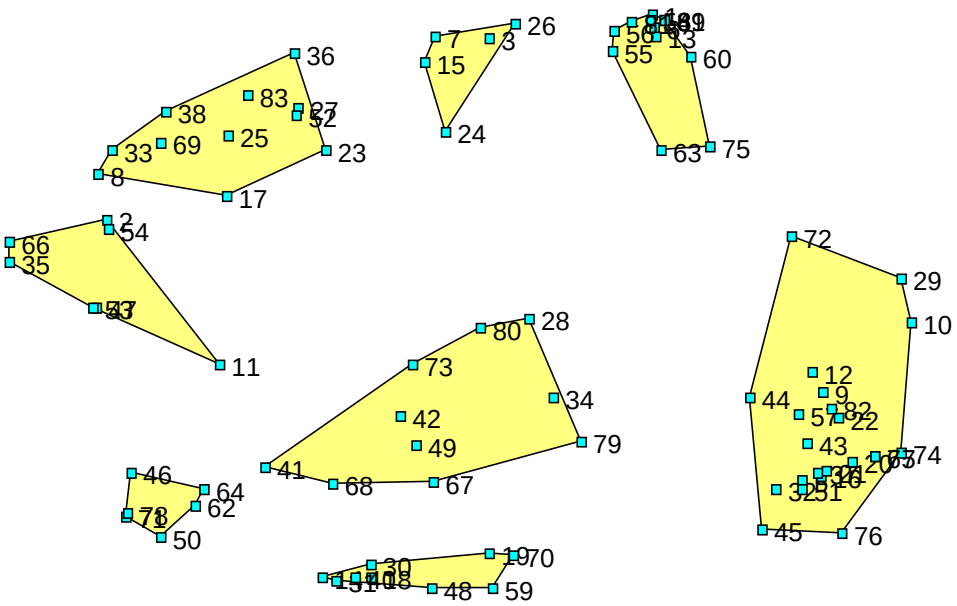


Figure 15. Point-cluster map for the OPEN project.

Each point on the map is shown with an accompanying statement identification number that enables one to identify the contents of the statements as shown in Table 1. We can also see the text of statements within each cluster, where the statements

are ordered in descending order by average importance across all participants as shown in Table 3.

While the map shows a considerable amount of detail, it is not in this form very useful as a graphic entity. In cartographic terms, it may be that the scale of the map is too detailed to be of much use in typical decision-making contexts.[8] To move to a higher level of abstraction or scale, we might drop off the detail of the points and include meaningful text labels for each of the clusters as shown in Figure 16.

How do we obtain the labels? First, the analysis uses an algorithm that mathematically determines the "best-fitting" label for each cluster based on the sort pile labels the participants developed. In fact, the algorithm enables one to select the top-10 best-fitting labels for each cluster in descending order of goodness-of-fit. Second, the participants examine the statements in each cluster and the top-10 labels, and from this information determine a cluster label that makes most sense.[9] Finally, the labeled cluster map can be graphed as shown in Figure 16.

This cluster map is often both interpretable and useful. in addition to its role as a summarizing and signifying device for the ideas of the group, it also often enables/encourages the participants to begin developing theories about the interrelationships among the idea clusters. For example, it is often sensible to "regionalize" the map by grouping clusters of clusters. In the map in Figure 16, it makes sense to group the four clusters on the top and left into one region that has to do with OPEN's relationship with the world external to it. The four clusters in this region -- Recruitment, Linkages, Outreach, and PR & Community Service -- describe the major areas of interface with this outside environment. On the other hand, the clusters on the lower and right side of the map pertain more to issues internal to OPEN -- the Programs & Events, Services, Communication, and Mentoring/Scholarships.

[8] Although this level of detail will be much more important when the organization actually begins to work on detailed design of a strategic plan with specific tasks or action steps). The point map is analogous to a detailed contour map or to a geological survey map - it doesn't give one a very good overall sense of the geography, but it is absolutely essential if you are going to hike through the area or are thinking of building there.

[9] In this demonstration project the author developed the final cluster labels by examining the statements in each cluster and looking at the top-10 labels. For instance, consider the cluster labeled 'Recruitment' in Figure 16. This cluster had as the top-10 labels (in descending order): Recruitment ideas, building membership, Scope, recruiting, Increase Diversity, Extend membership, Members-who, Outreach-Non-Evals, Membership, Member recruitment. After reviewing these labels and the statements in that cluster, the term 'Recruitment' was selected as the final label that best summarized the ideas.

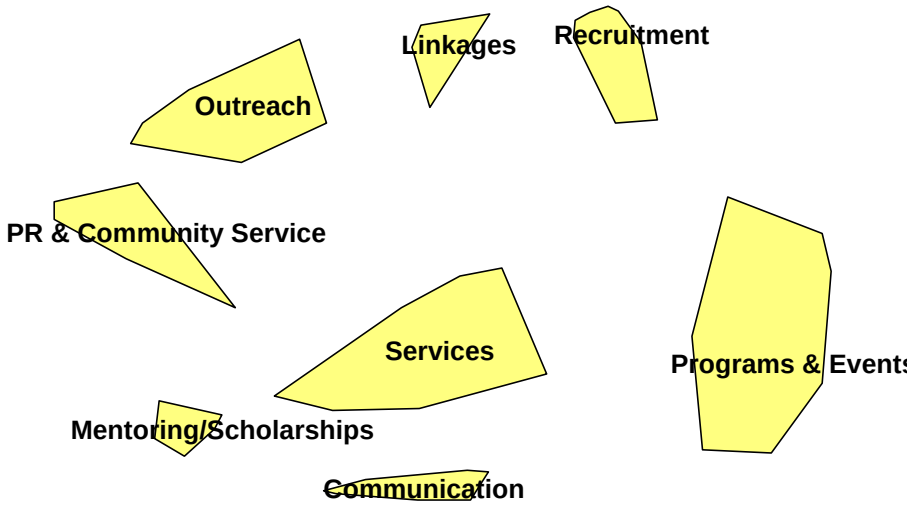


Figure 16. Labeled cluster map for the OPEN project.

The labeled cluster map in Figure 16 can be considered the "base" in this relational map. Just as with the individual points, distance between clusters is interpretable.[10] Typically, we use this base as the foundation for displaying or representing other information. For instance, consider the map in Figure 17.

[10] Dimensional orientation is irrelevant -- the map does not know north from south or east from west. once could rotate the map and/or flip it horizontally or vertically and the points/clusters would remain in the same relational configuration. The shape of the cluster is also irrelevant -- it is determined solely by joining the outer points in each cluster. however, the size of the cluster does have some interpretability. In general larger clusters are "broader" in meaning and smaller ones are more 'narrowly" defined.

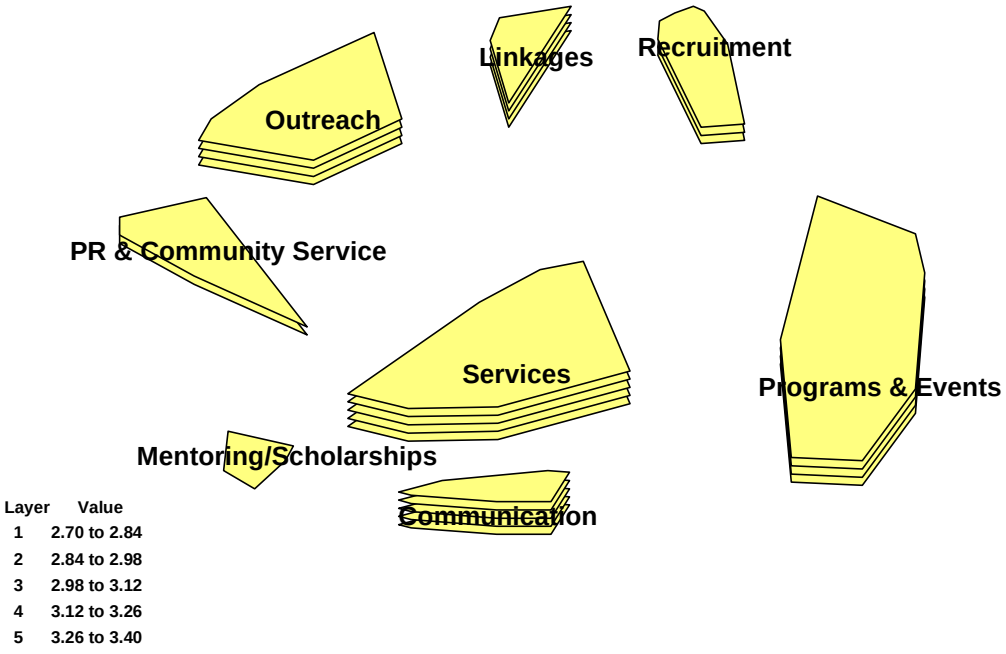


Figure 17. Cluster rating map showing average importance across all participants.

In this map, we use layers on the base cluster shapes to represent a separate variable, in this case, the average importance rating across all participants.[11] This type of map is analogous to the map of crime rates in the United States that is shown in Figure 4 except that in this map the base is conceptual rather than geographical.

As with any geographical mapping system, it is often useful to compare two different variables as represented on a relational base. This is illustrated in Figure 18 which shows on the left side the pattern of ratings for participants who were government employees and on the right side the pattern of ratings for participants who were not government employees. This might be analogous to two maps of the United states that show felonies and misdemeanors.

The problem with these types of maps is that it is difficult to compare the information visually. Such a comparison requires the eyes to flit back and forth between the two figures, noting what it high in one and how high it is in the other.

[11] The map also has a legend that shows that the cluster averages range from a low of 2.70 to a high of 3.40, even though the rating was measured on a 1-to-5 scale. This narrow range often surprises people who are tempted to conclude that it means there is essentially no difference between the highest and lowest rated cluster. However, it is important to remember that the cluster averages are aggregated twice: for each statement the average is determined across participants and for each cluster, the average is determined across statements in the cluster. Because of this, it is expected that the range would be considerably restricted. it also means that even relatively small differences between cluster averages are more likely to be interpretable than might appear at first glance.

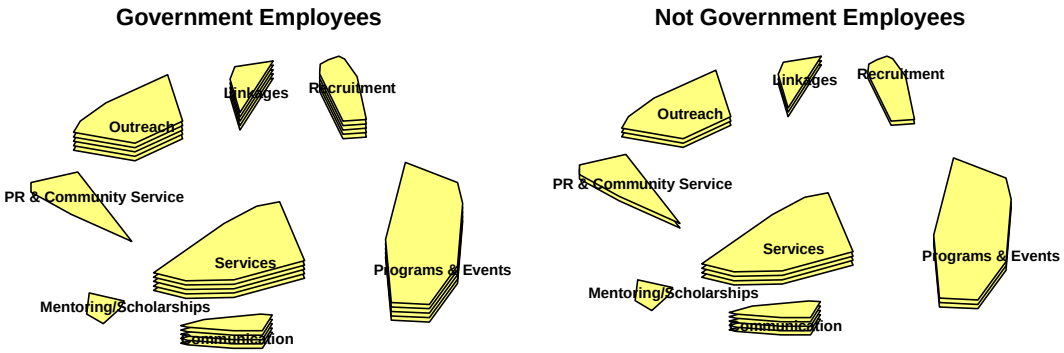


Figure 18. Comparison of the patterns of ratings of government employee participants and non-government employee participants.

Another way to accomplish the same "pattern match" comparison more effectively is to array the two patterns side-by-side in a "ladder graph" form as shown in Figure 19. The ladder graph has two vertical axes, one for each group. Each cluster from the map is represented as its own line on the ladder and is listed vertically as a label. The ordering of the labels from top-to-bottom depicts the rank ordering of the importance of the clusters and links to the order in which each cluster's line hits the axis on that side. The point where each line hits the axis depicts the interval-level average of the importance ratings. The Pearson Product Moment Correlation (r) is an estimate of the overall pattern match and is shown at the bottom of the figure. This type of figure is called a "ladder graph" because, when there is strong agreement between the groups, the lines tend to be horizontal and look like a funny type of multi-color uneven-step ladder.

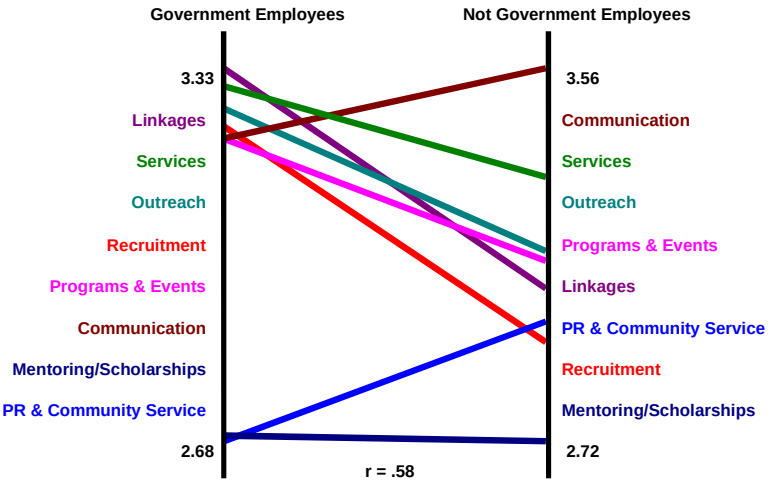


Figure 19. A ladder graph representation of the pattern match data in Figure 18.

The pattern matching ladder graph is especially useful for examining visually the degree to which two vertical pattern representations on a concept map base are similar or different from each other. For example, the patterns in Figure 19 suggest that while Linkages is the most important cluster for government employees, it is slightly below half-way down in importance for non-government employees. We can also see in the figure that government employees have almost dichotomous or bimodal ratings -- Mentoring/Scholarships and PR & Community Service are rated almost equally low and everything else is almost equally high.

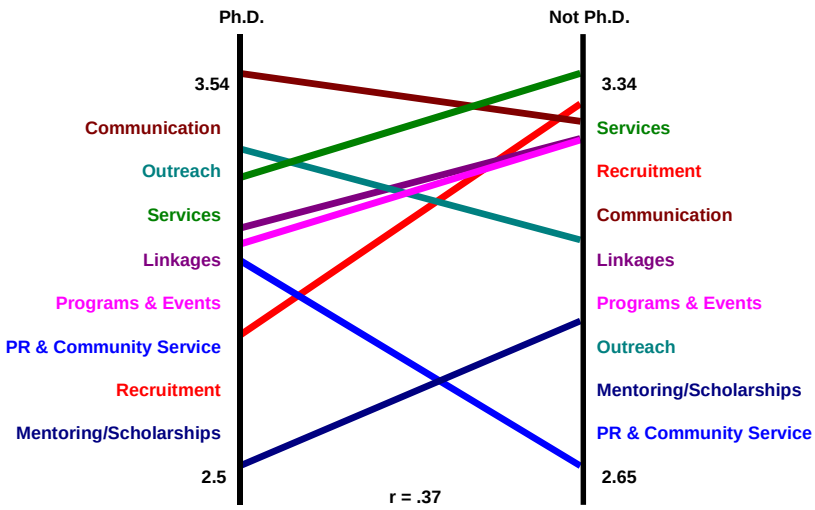


Figure 20. Pattern match of participants having a Ph.D. versus those who do not.

Figure 20 Shows the pattern match of respondents with a Ph.D. compared to those without one. Several disconnects seem interesting. Non-Ph.D.s seem to rate Recruitment and possibly Mentoring/Scholarships more highly than Ph.D.s do. On the other hand, they rate PR & Community Service lower, and it is their lowest category.

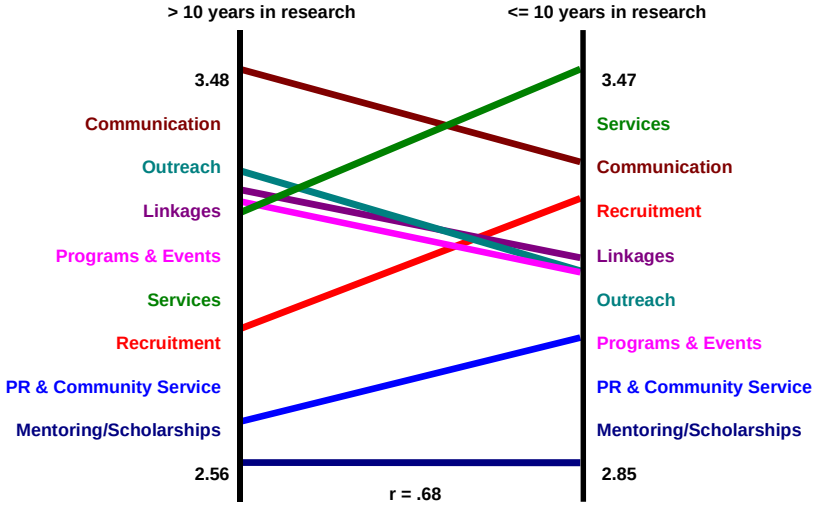


Figure 21. Pattern match of those with more versus less than ten years of experience with research.

In Figure 21, we see that the less experienced researchers seem more interested in both Services and Recruitment than their more experienced counterparts. This makes sense. They are more likely to be in the process of establishing their careers and in need of assistance/guidance (e.g., Services) and interested in networking and making contact with others who are like them (e.g., Recruitment).

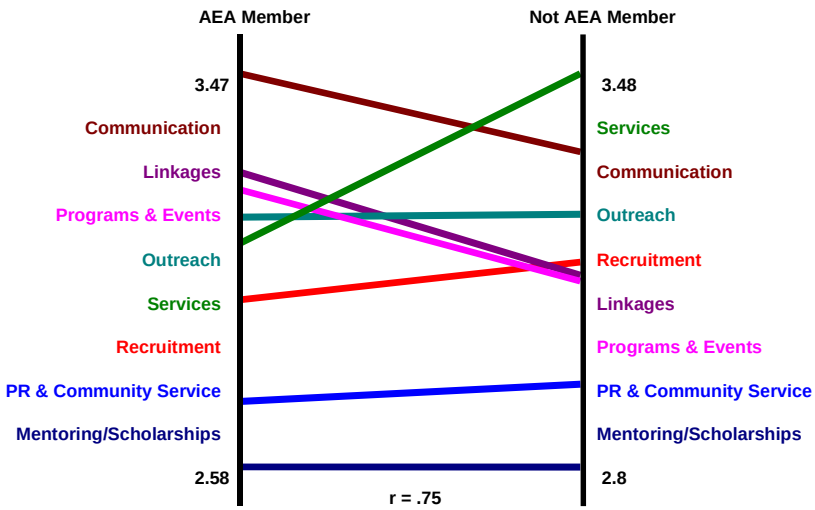


Figure 22. Pattern match of AEA versus non-AEA Members.

The national association of evaluators in the United States is the American Evaluation Association (AEA). It stands to reason the OPEN members would be interested in participating in this national association in addition to OPEN. Figure 22 shows that OPEN members who are *not* members of AEA view Services as the most important cluster, most likely because they rely on OPEN as their primary or sole source for such services while AEA members have at least that alternative.

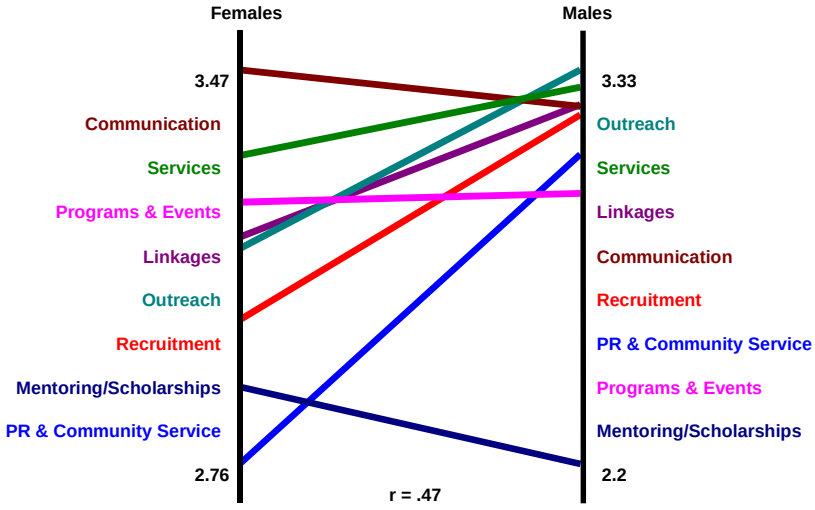


Figure 23. Pattern match of females versus males.

Gender was also compared as shown in Figure 23. Here, the most salient feature appears to be the degree to which males rated Mentoring/Scholarships lower than all other categories -- their ratings are almost bimodal in this sense. Can this be reflective of a general tendency of males to be less interested in mentoring and sponsoring activities?

Pattern matches can be portrayed either relatively of absolutely, and there are distinct advantages and disadvantages to each. For instance, consider the two pattern matches shown in Figure 24. Both ladder graphs portray exactly the same pattern matching data. Both compare the average importance of participants living in Portland, Oregon with those not living in Portland. The only difference between them is that the ladder graph on the left uses the actual minimum and maximum average

ratings for each group to determine the axis high and low values whereas the graph on the right fixes the axes for both groups to the same high and low values.

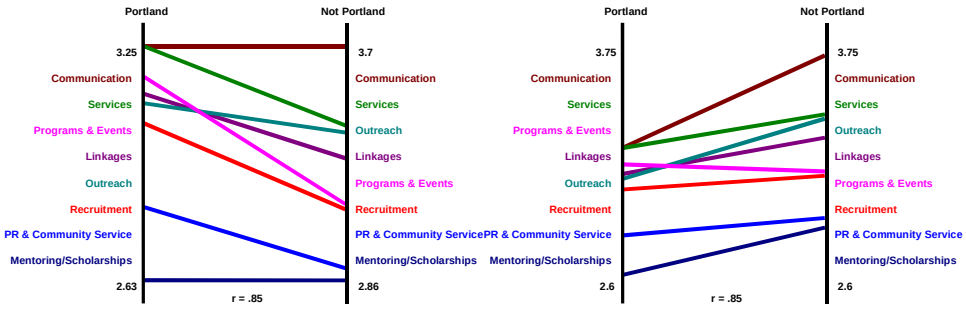


Figure 24. A relative (left) and absolute (right) pattern match for the same comparison.

Both pattern matches show that Communications is important to both groups, but the absolute match shows more clearly that it is more important to those outside of Portland. On the other hand, the relative pattern match shows that, relatively speaking, Programs & Events is rated higher by those in Portland than by those outside it (perhaps because those farther from Portland find the events predominantly in Portland less accessible to them).

What can we make of these results? We need to recognize, of course, that they are based on very few respondents. This project was undertaken as a demonstration, not as a more formal evaluation. Inferences from so small and unrepresentative a sample should be made very cautiously. In addition, all interpretations made here are from the perspective of the author who was also the facilitator of the project. There was brief opportunity to engage with OPEN members at their conference and present the results, but no attempt comparable to what we typically do in concept mapping was made to engage the groups in the process of interpretation.

On the other hand, we can look at these maps and matches in the spirit of "suggestive" devices designed to stimulate members of OPEN to think about their organization and its issues in different ways. With this in mind, and not taking any of the results too seriously, a few inferences do "suggest" themselves. It seems clear from the map that there is an internal-external division in how people view OPEN and its issues. The most salient internal issues relate to the categories Services, Programs & Events, Communications, and Mentoring/Scholarships. The external issues are categorized as Recruitment, Linkages, Outreach, and PR & Community Service. It's interesting in reference to the maps shown in Figure 16 and Figure 17 that Recruitment (external) is located near Program & Events (internal) while PR & Community Service (external) is closer to Services and Mentoring/Scholarships (internal). In terms of planning, this may suggest that the organization look for internal-external synergies in these areas. For instance, this may suggest that Programs & Events be viewed as a major device to help achieve Recruitment. Conversely, one might conclude that more Recruitment is important in order to enhance Programs & Events or offer a greater variety of these. Analogously, one might look at the confluence of PR & Community Service (external), Services (internal) and Mentoring/Scholarships (internal) and conclude that the Mentoring/Scholarships category might very well be a "bridge" area -- it can help enhance PR & Community Service for OPEN while at the same time providing a Service to OPEN members.

The other major interpretation from this brief pilot project is related to the pattern matches. These suggest that OPEN has a number of important constituencies who have some discernable differences in their preferences. They look to the organization

for different things, and participate in different ways. This is no surprise given the diversity of members, but the pattern matches do help suggest which groups have which preferences.

Finally, it's important to recognize the stated purpose for this pilot study (outside of its obvious use as a demonstration of concept mapping). The focus of the study was on strategic planning for OPEN -- what OPEN should do over the next five years. We would hope that the map -- like all maps -- would have some utility in guiding the organization, in this case, in their strategic planning efforts. How could it be used in this way? First, the maps act as a communications device. They summarize an enormous amount of information into eight categories, providing a common language that can shape subsequent planning. That is, they help to organize and simplify the discussion. Second, the maps provide an implicit structure that can be used to organize strategic planning efforts. For instance, OPEN might decide to subdivide the next phase of its strategic planning effort by forming several committees or task forces. In lieu of this map, they would need to create these intuitively. The map, however, suggests eight categories that might form the basis for several committees. If they only want a few committees, they might use the internal-external distinction. If they want more, they could have one per cluster or group adjacent clusters (e.g., fold Mentoring/Scholarships into the adjacent Services cluster and form a Services committee). Each such committee already has a good head start. In their designated clusters they have a set of issues (i.e., the statements as shown in Table 3) that have been rated for importance. They might start with the ones rated most important by all participants and begin to develop potential action statements, deadlines, assignments, and so on[12]. Or, the group might prefer to work as a whole and might look across the entire map to find the statements that were rated most important across all participants as shown in Table 4. For these they might then decide on actions and the details associated with implementing them. These are simply initial suggestions for how the organization might use the maps to guide their strategic planning. Potential uses are limited primarily by the creativity of the organization.

It should be clear that concept maps, like their geographic counterparts, are intended to provide guidance, a sense of direction. Instead of geographical terrain, the concept maps attempt to describe the conceptual and semantic terrain. With this case study in mind, we turn now to look at what cartographers are saying about mapping, to see how their insights and experiences might illuminate the evaluator's endeavors.

The Cartography of Concept Mapping

A central premise of this paper is that evaluators can, both figuratively and *literally*, be cartographers of the social reality within which they work. When understood in the broader classification described above, it should be clear that the evaluator-cartographer is not meant to be limited to geographic mapping (such as geographic information systems). In fact, geographic mapping may in the end be the lesser application of the evaluator-cartographer. Relational maps, especially ones that

[12] Concept Systems Incorporated, the producers of The Concept System, have a number of post-mapping modules that help participants take the results of maps and use them for various purposes. These modules, termed the Concept System Application Suite, include (among many): a program for strategic or action planning that enables an organization to develop tasks for any statement/clusters, set deadlines, assign resources, set costs, and so on; and a program for evaluation or performance measurement that enables the development of multiple measures for each statement/cluster, specification of targets or objectives, and ongoing monitoring over time, analysis, and reporting.

involve additional representation of data, are potentially valuable both for prospective, formative evaluation and retrospective, summative work.

If one accepts the premise that mapping has a useful place in evaluation, and that the type of non-geographical mapping that might be undertaken by evaluators is conceptually related to traditional geographic mapping, it makes sense that we might examine how contemporary cartographers think about quality in their work to see if there may be some useful lessons and/or warnings there for evaluator-cartographers.

The naïve view of most laypersons is that geographic maps -- those that we might call geographically isomorphic -- more-or-less correspond to external reality. A map, viewed from that perspective, can be judged in terms of its *accuracy*, the degree to which it correctly reflects geographical reality.

Those of us who have been engaged in evaluation work are familiar with this perspective with respect to evaluation. Just as with maps, our evaluations are presumed by many laypersons to be isomorphic with reality. An evaluation in this perspective can be judged in terms of its validity or the degree to which it "correctly" represents reality. In Harley's words:

> The usual perception of the nature of maps is that they are a mirror, a graphical representation, of some aspect of the real world. …the role of a map is to present a factual statement about geographic reality. (Harley, 1990)

The positivist and naïve realist perspectives in evaluation have been successfully challenged by a variety of post-positivist stances that include critical realism, emergent realism, constructivism, and interpretivism, among others. All of these perspectives share the premise that reality is not an immediately sensible phenomenon that can easily be represented acontextually. All are critical of our ability to observe and measure "accurately" in the naïve positivist or realist sense. All of them view the political and interpersonal environment as critical components in evaluation contexts.

It is somewhat heartening as an evaluator to read as an uninitiated outsider some of the recent literature on cartography. Many of the themes raised there, and many of the conclusions that result, have direct analogues in evaluation. The cartographers, like evaluators, have struggled with the positivist and naïve realist paradigms. They, like we, have recognized the critical importance of context, the frailty of measurement and observation, the degree to which our work is constructed based on our perspectives and experiences.

In this section some of the current thinking in cartography is reviewed. No attempt is made to represent accurately the full range of current debate among cartographers. It is almost certainly as divided and contentious as current debate among evaluators. This discussion also does not pretend to represent the most discerning reaches of academic discourse in cartography. The primary sources used here are readily available in the popular press and can be easily located in most large bookstores.

Although several sources are included, this discussion draws heavily from the fascinating work by Denis Wood entitled The Power of Maps (Wood, 1992). Wood clearly comes from a constructivist/interpretivist perspective in cartography that challenges even the assumptions of post-positivist critical realism. His insights are challenging and have great potential value for evaluator-cartographers, whether one agrees with him or not. The discussion that follows is based on a set of themes, each

of which is stated directly from the contents of Wood's book.[13] Within each theme is some brief discussion of the theme within cartography along with consideration of the implication of the theme for the concept mapping case study in particular and concept mapping in general.

Every Map Has an Author, a Subject, a Theme

> …maps, all maps, inevitably, unavoidably, necessarily embody their authors' prejudices, biases and partialities … There can be no description of the world not shackled … by these and other attributes of the describer. Even to point is always to point…*somewhere*; and this not only marks a place but makes it the subject of the particular attention that pointed there instead of … *somewhere else*. The one who points: author, mapmaker; the place pointed: subject, location; the particular attention: the aspect attended to, the theme -- nothing more is involved (and nothing *less*) in any map. (Wood, 1992, p. 24)

Who is the "author" of a concept map? We usually describe three types of people who are directly involved in developing a map -- initiator(s), facilitator(s) and participants. The initiator(s) is the person(s) who called for the map, who set the process in motion. In most mapping contexts, the initiator is someone from a client group or organization who contacts the facilitator(s) who will oversee the process itself. The initiator is usually fairly powerful within the context of the map interest. They might be the director or managers in an organization. By selecting a concept mapping process itself, they are "pointing" in a sense, choosing one method or approach over another. But they have a more important influence in shaping the project because they are usually the key persons working with the facilitator to decide on the focus for the map, the demographics and ratings, the schedule and who will be selected to participate. The facilitator also plays a key role in shaping the nature of the map. As the person most likely to be knowledgeable about the mapping process and how it works, the facilitator has numerous opportunities to shape the content. For instance, the facilitator plays a key determinative role in:

- shaping the planning (selection of focus, schedule, participants, etc.)
- shaping the content of the brainstormed statements through direct management of the group during brainstorming and any editing of the statement set
- shaping the ratings by helping phrase the rating instructions
- shaping the analysis through choices in analysis parameters
- shaping the results, especially through the selection of the number of clusters and in the rearrangement of any points or cluster boundaries on the map
- shaping the interpretation, through what is presented to the participants, in what order, and through management of the interpretation discussion

Obviously, it's also important to recognize that the participants shape the content of the map. In one sense, they are considered to be the primary authors. However, it

[13] The contents alone are a stimulating read. The notion of stating contents headings in so direct and accessible a manner adds greatly to the readability and impact of the volume. However, the themes used here are only a partial representation of the contents. Many of the themes that are omitted relate to the iconographic nature of many geographical maps -- a feature not shared by the type of concept mapping described here.

should be clear that their contribution is shaped significantly by the initiator(s) and facilitator(s).

One of the most dangerous and potentially deceptive aspects of concept mapping is related to the apparent and real roles of the facilitator. On the surface to many participants and stakeholders, the facilitator appears to be a disinterested one who objectively helps the group surface their content. In reality, the facilitator plays a major determinative role, influencing in many subtle ways. However, it may not be clear in what directions or for what interests the facilitator works. As facilitator of this project I was not a member of OPEN, only knew professionally a few members (including the initiators) and had no real interests of my own other than wanting to do as nice a demonstration of the method as possible (which is not to say this isn't a powerful influencer). But, I am a white, middle-age professor from an ivy league school and with extensive experience in the evaluation profession. These factors certainly influence decisions I make or encourage others to make as part of concept mapping. I hasten to add that I don't think this is a problem in itself -- any facilitator will have biases and unique perspectives that they bring to a project. The danger is not in the bias, it is in not recognizing that there is that influence. To pretend that the facilitator is an objective outsider who only facilitates and does not influence the results which, instead, stem entirely from objective and rigorous statistical and computer-based processes -- now *that* is a potential danger.

In the OPEN case study, I was the only facilitator. I was contacted by the current President of OPEN who knew of my mapping work and asked me to be the keynote speaker for their first conference. In our early discussions I suggested that we do an actual concept mapping project for OPEN that could then be incorporated into my talk. Subsequently another associate of the President's joined our discussions because he was both a member of OPEN and experienced with the concept mapping methodology. Therefore, in this context we had one facilitator and two initiators from the client organization. All of the planning decisions for the project were made by the three of us. We decided to do the process (up to interpretation of the maps) using the web version of The Concept System. We decided on the focus statement for brainstorming and we decided to have all OPEN members on their e-mail list (evidently the vast majority of members have e-mail, and OPEN uses it routinely for member communications, although it's not clear how much/often members use it). All e-mail communications about the project came to members directly from the OPEN President, although I actually did all the management of the technology and process over the web.

There were several points at which I or the initiators influenced the process, in addition to the decisions described above. During the brainstorming period very few statements were being generated over the web. A few days before the brainstorming was due to close, I sent this e-mail to the initiators:

> It's Tuesday afternoon and we only have 24 brainstormed statements! I'm shutting this thing off this weekend and I REALLY, REALLY hope you guys can drum up a bit more involvement than that. I'm going to spend a fair amount of this Sunday generating 300 or so datasets for your participants and it wouldn't hardly be worth it unless we get 70-80 statements. Of course, you two can "pad" the list when you do the editing this weekend (and you may have to), but it would be so much nicer if others participated.

Of course, this was an artificial example among friends and done entirely for purposes of demonstration. But it's clear from the message that I'm concerned both about the low level of participation and the degree to which it might reflect poorly on my methodology (not to mention the fear that I would not obtain a good enough result to present and would have to come up with another example!). My initiator buddy wrote me back:

> only 30 statements: What a bummer! Especially since I did 15 of them already. So yes, I will start padding --also will call friends and solicit their statements and add them as well.

He evidently did a great job because, in the end, we had 83 statements. Was I concerned about this participation? Not in this case because it was clear only a demonstration. Could stuff like this happen in "real" projects? Absolutely. The problem is not particularly that this kind of thing occurs -- it's that it doesn't get presented. The impressive-looking and authoritative maps that are presented look as though they are based on extensive participation. This issue manifests itself in different ways depending on how the brainstormed statements are gathered. For instance, when brainstorming is done over the web as in this case study, it is virtually impossible to tell how broad-based the participation is because it is anonymous[14]. Different and equally vexing problems occur in a face-to-face brainstorming session.

When brainstorming was shut down, I instructed the initiators to "edit" the brainstormed statement set. Here's part of the e-mail interaction:

> **FROM Initiator:**
>
> Just to make certain I'm clear about the goal-we want statements that are conceptually different from each other as opposed to reiterations of similar themes? What about statements that are more detailed aspects of broader concepts (e.g., "Hold more events outside the Portland area" vs. "Schedule meetings in Salem (or Eugene, or Bend...")? Is it legitimate to include all of those statements about the "meeting venue" concept, or should we cull out the more detailed statements? Or should we include the more detailed statements, and leave out the broader concept?
>
> **My Response:**

> Technically, as long as the statement completes the focus prompt it is legitimate. However, you want to use some judgment when you revise. if someone says, "hold more meetings outside Portland" I guess I would be inclined to leave it alone. However, if you get a ton of statements that say "hold a meeting in Portland", "hold a meeting in Eugene", "hold a meeting in Podunck", I would probably collapse these to "hold meetings in a variety of locations throughout the state."

Here, both the facilitator and the initiators are influencing the map results directly in the numerous editing choices. One might argue that we are trying to be fairly even-handed in our judgments. However, if you were a member from an area of the state that had never had an OPEN meeting, having your statement lumped in with others who have had prior meetings in their areas might legitimately lead you to feel we were biased against you in our editing judgment.

In the end, the statements themselves should reflect legitimate content from the participants. Thus, in concept mapping there are multiple "authors" who influence the subject and theme of a map in numerous often unconscious ways. Like our cartographer colleagues, the evaluator-facilitator in this context often tries to distance themself from the result. We are only facilitators. It's the map that is an objective

[14] Actually, for those who know about how the web works, you should recognize that while we don't require login for brainstorming, every web server can trace every hit to the web back to a specific IP number. In many organizations, IP numbers are assigned permanently to specific machines, which means we could identify which machines hit the brainstorming. Of course, we don't know who is on that machine, but in many companies a machine is assigned to an individual, so we might infer who it was that hit the brainstorming site. Of course, the log won't tell us what statement was brainstormed in what hit, although it's possible to reconstruct the sequence. So, theoretically, it is possible that one could identify who brainstormed what statement even with "anonymous" brainstorming. I wouldn't want to try figuring it out, however.

reflection of reality. The cartographers remind us that that isn't even true in most geographic mapping! Why would we expect it to be in concept mapping?

Every Map Shows This…But Not That

> …the cartographer makes a selection, classifies, standardizes; he undertakes intellectual and graphical simplifications and combinations; he emphasizes, enlarges, subdues or suppresses visual phenomena according to their significance to the map. In short, he generalizes, standardizes, and makes selections and he recognizes the many elements which interfere with one another, lie in opposition and overlap, thus coordinating the content to clarify geographical patterns of the region. (Imhoff, E., 1982)

What is the analogue in the type of concept mapping described here? Perhaps it is simply the choice of this form of concept mapping, this specific methodology, that, like a choice of projection method in geographic mapping is a key selective factor (although it is difficult to see how that choice itself determines the content of a map). In another sense, the methodology itself -- the algorithms used, the wording of the focus statement -- limits the nature of the resulting map. Perhaps most obvious, the concept map may be most limited in its perspective by the selection of the participants and by the contextual circumstances that might lead them to participate actively and offer what they really think and believe. In this sense, anonymous participation rather than face-to-face mapping may open up the process and help to mitigate the narrowing of ideas. On the other hand, as this case study shows, anonymous participation has its own problems.

One of the most delimiting aspects of concept mapping is the brainstorming focus statement. It determines what content is allowed and what is excluded. Here, again, is the one we used in the case study:

> *One specific thing I think the Oregon Program Evaluators Network should do over the next five years is…*

At first glance, this isn't a very limiting statement. But let's examine it a bit more critically. It limits the statement to a five-year horizon. That's pretty broad, but it doesn't encompass more than a portion of the time span of most members' careers as evaluators. The focus calls for things that can be done -- for actions. It is neutral on whether they should ever be fully accomplished, although that it suggested. The focus is also neutral about why the suggestions should be done. To improve OPEN? To help members? To help control the evaluation profession in Oregon? Perhaps most important, there are numerous potentially critical things that are outside the scope of the focus. The focus doesn't direct you to brainstorm about the current problems OPEN has, or the things you like about the organization. In this sense, the focus makes a choice, one that influences everything that follows in the process.

Another place where choices are made is in the selection of the number of clusters on the map. There is no easy way to do this mathematically, and it's not clear that it would be desirable to do so even if there were. This is also not an easy thing to accomplish in a group process -- most groups have neither the interest or to examine multiple cluster solutions and make a determination of which seems best. This judgment is usually left to the facilitator. In this case study, I followed the standard procedure for selecting clusters (see above) and the process was fairly straightforward. However, in other mapping processes there are times when the maps are not as clear as in this case study, or where there are multiple cluster solutions that would be of value. This is a bit like the cartographer who recognizes that drawing a map at one scale has value for one purpose while drawing it at another

has value for some other. It is not that one is "right" or "wrong", "better" or "worse" in some abstract sense. It is that the context is in part determinative of the value of such choices.

Maps Construct -- Not Reproduce -- the World

The naïve view is that a geographical map is a window on reality.

> Were it not reality, why then it would just be…*opinion*, somebody's *idea* of where your property began and ended, a good *guess* at where the border was, a *notion* of the location of the hundred-year flood line, but not the flood line itself. What is elided in this way is precisely the social construction of the property line, the social construction of the border, the social construction of the hundred-year flood line, which -- like everything else we map -- is not a line you can see, not a high water mark drawn in mud on a wall or in debris along a bank, but no more than a more-or-less careful extrapolation from a statistical storm to a whorl of contour lines. As long as the map is accepted as a window on the world, these lines must be accepted as a representing things in it with the ontological status of streams and hills. (Wood, 1992, p. 18-19)

But this is more than simply a philosophical conundrum, an abstract academic argument. As in evaluation, one's position on this issue has a dramatic effect in real-world consequences:

> But no sooner are maps acknowledged as social constructions than their contingent, their conditional, their…*arbitrary* character is unveiled. Suddenly the things represented by these lines are opened to discussion and debate, the *interest* in them of owner, state, insurance company is made apparent. (Wood, 1992, p. 19)

In concept mapping, the constructivist nature is even more apparent perhaps than with geographical maps. But, as we utilize more and more sophisticated equipment and rely on more advanced mathematical algorithms, the temptation increases to see the results of all this sophistication as something accurate, a valid representation of an implicit cognitive world.

This is a perfectly reasonable impulse. And, by making this argument we don't presume that we should give up on the goal of making as accurate a map as possible. We won't contend that a random concept map, thrown together with randomly generated mathematical data is a reasonable goal any more than Wood is likely to contend that random sketches might qualify as geographically useful guides. The point -- for both geographical and concept maps -- is to recognize that no single map can possibly *reproduce* reality. Any single map is but one of an infinity of other potential reasonable and accurate reflections of some aspect of reality. Rather than being a constraint, this recognition actually acts as a catalyst. For the map user, the map becomes one view, suggests other views, should be critically viewed. The question that every map raises, at least implicitly, is "Why was this question asked?" or "Why was this approach taken?" or "Why was this perspective used?"

> Not that accuracy is not worth achieving, but it was never really the issue, only the cover. It is not precision that is at stake, but precision with respect to what? What is the significance of getting the area of a state to a square millimeter when we can't count its population? Who cares if we can fix the location of Trump's Taj Mahal with centimeter accuracy when what would be interesting would be the dollar value of the flows from the communities in which its profits originate? What is the point of worrying about the generalization of roads on a transportation map when what is required are bus routes? Each of these windows is socially selected, the view through them socially constrained no matter how transparent the glass, the accuracy not in doubt, just…*not an issue*. (Wood, 1992, p. 21)

The OPEN map is clearly a construction. It is founded on an unrepresentative sample of OPEN members and based on algorithms that inevitably introduce distortions.[15] It addresses a very broad topic -- what OPEN might do as an organization over the next five years -- from the perspectives of a few self-selected individuals. These participants may have a longer-term interest in OPEN than most. They are likely to differ in numerous ways from members who chose not to participate. It raises a question about the organization within a specific time frame. Others might be interested in different questions, different time frames. This doesn't mean the map is not "accurate," only that it has a point of view from which it is constructed. We might obtain any number of other maps that are relevant, interpretable, or useful, from the same participants or from others. We might find that, even if we replicate the study on different OPEN members, the results are substantially unchanged. Or not. That issue is ultimately empirical. But the choice of perspective, the questions asked, the experience base, are reflections of decisions made consciously or otherwise, in the course of producing this map. Even a larger sample (or complete participation from every member of the population of interest) *would do nothing to diminish the fact that the map is inherently a construction based on an infinity of choices*.

Maps Make the Past and Future Present

> The world we take for granted -- the real world -- is made like this, out of the accumulated thought and labor of the past. It is presented to us on the platter of the map, *presented*, that is, *made present*, so that whatever is invisible, unattainable, erasable past or future can become part of our living…*now*…*here* (Wood, 1992, p. 7)

> This is what it means to use a map. It may look like wayfinding or a legal action over property or an analysis of the causes of cancer, but always it is this incorporation into the here and now of actions carried out in the past. This is no less true when those actions are carried out…entirely in our heads: the maps we make in our minds embody experience exactly as paper maps do, accumulated as we have made our way through the world in the activity of our living. (Wood, 1992, p. 14)

Here, the link between geographic and conceptual maps is almost explicitly drawn. All concept maps, like geographical ones, incorporate our past into an imperfect agglomeration of our experiences and thoughts. The OPEN concept map represents the past through the experiences of those who participated. It anticipates the future in the sense that it is focused on things OPEN *could* do. Thus, this concept map, like every other, is historically indexed, a snapshot in time that presents the past and the future at one moment.

Maps Link the Territory with What Comes with It

> It is this ability *to link the territory with what comes with it* that has made maps so valuable to so many for so long. Maps link the territory with taxes, with military service or a certain rate of precipitation, with the likelihood that an earthquake will strike or a flood will rise… (Wood, 1992, p. 10)

15 For instance, the multidimensional scaling is restricted to two dimensions when, in reality, we have no idea how many dimensions might be needed to represent the information accurately.

Analogously, it is the ability of a concept map to link the concepts or ideas with action that makes it so valuable. In a planning endeavor, we often use the clusters as the basis for action. Statements in the cluster become the basis for action statements that have specific persons assigned to carry them out, resources attached, deadlines set. In evaluation contexts, we often use the map as the basis for measurement construction. The cluster acts as the measurement construct. The statements are potential or actual measurement items, operationalizations of the construct. In both contexts, the map provides a vehicle or base for representation of the consequences or results. it is the framework for showing progress on the strategic plan or change in the measurement.

In thinking about geographic maps, the cartographer has reference to the implicit geographical maps we carry around in our minds -- the mental maps -- of the geographical landscape that we traverse. Analogously we carry around mental maps that extend beyond simple geography, for our mental maps incorporate various modes of our experience, the sights, smells, sounds and, especially the language and meaning that we integrate with that experience.

> Of course a mental map is … clearly related to the way we use paper maps to make decisions. Certainly the similarities increase once we begin to externalize these maps, to share them with each other. "What? Why would you go that way?" "Because it's shorter." "No, No, it's shorter if you take St. Mary's to Lassiter Mill -- " "Oh, and then go out Six forks to Sandy Forks?" "Yeah." Here the maps, in separate heads, are being consulted almost as if they were paper maps open on the table, linking knowledge individually constructed in the past to a shared living unfolding in the present. (Wood, 1992, p. 15-16)

Concept maps also enable individuals to negotiate through a terrain, in this case a cognitive terrain of semantic relationships and personal experiences. The discussion of a concept map is a useful underpinning for the exploration of shared meaning and consensus or the identification of diversity and divergence of view. In the OPEN case study, this discussion has not (and may never) occur. There was some brief discussion of the maps produced here at the conference. The major purpose of constructing these maps -- to illustrate the mapping process on an example of relevance to OPEN members -- was accomplished at the conference itself. Nevertheless, OPEN could decide to embark on strategic planning using the concept map produced here (and, one would hope, augmenting it with broader involvement and input).

The Interest the Map Serves is Masked

Why do so many people bristle at the idea that every map is biased, perspective-laden, a necessary distortion?

> It is because the map is powerful precisely to the extent that this author…disappears, for it is only to the extent that this author escapes notice that the real world the map struggles to bring into being is enabled to materialize (that is, to be taken for the world). As long as the author -- and the interest he or she unfailingly embodies -- is in plain view, it is hard to overlook him, hard to see around her, to the world described, hard to see it…as the world. …As author -- and interest -- become marginalized (or done away with altogether), the represented world is enabled to…fill our vision. Soon enough we have forgotten this is a picture someone has arranged for us (chopped and manipulated, selected and coded). Soon enough…it is the world, it is real, it is…reality. …When the authors have rendered themselves transparent, suddenly we have…*fresh respect*. And it is astonishing how easily this happens, how readily we take for granted -- as natural -- what is never more than the social construction of a map. All that is required is the disappearance of the author, the invisibility of the interest. (Wood, 1992, 70 - 71).

Concept mapping benefits from the illusion of objectivity, the appearance that there is no one who is driving the machine, that it is a dispassionate and scientific algorithm that determines the arrangement. But there are many places where the person behind the algorithm enters in, and even more where the participants themselves shape the result.

> Again and again we have seen a similar vagueness of content and form, a similar *diffusion* of ends and means. The map will show everything (and therefore claim innocence about the choice of anything) and will show it as it is (ignoring the "white lies" the map must tell in order to be accurate and truthful). The map will be seen to serve so many purposes that none can predominate, or its means will be so widely spread in so many social institutions that it can be claimed by none. Responsibility for the map will be shuffled off onto layered and competing interest groups, or its authorship will be shown to be fragmented among so many specialists as to be impossible to establish. Lying and lost, vague and confused, the map will *therefore* show the world the way it really is. (Wood, 1992, p. 73)

I find this to be a particularly apt insight. There is something compelling about concept maps. Despite the countless decisions and judgments that go into them, the multiple perspectives and distortions, it is hard to look at a final map and refrain from treating it as "accurate" in some sense. Our species has evolved an extremely powerful desire to "make sense" of our environment. After all, we're able to do elaborate interpretations even of random ink blots! A concept map shows a picture that always appears to be very orderly and sensible. The computer generates very crisp looking polygons to enclose groups of ideas. The layers suggest relative importance and other characteristics. In the end, the map is often just plain *compelling*. We have to remind ourselves of its frailty, its uniqueness. We have to be honest with ourselves and the participants about how it masks various interests and encompasses biases.

The Interest the Map Serves Can Be Yours

Some might see the perspective offered here on concept mapping and be depressed or discouraged by it. To recognize that an apparently very powerful tool is (like all other tools, I hasten to add) fallible and imperfect is sometimes difficult to take. I'm sure that the recognition in the cartography community that this is so for geographic maps has been a source of some contention and consternation there for the past few decades.

Once we accept the inherent perspective-ladenness of every map, once we see that each map serves interests, does the map itself lose its value?

> Once the map is accepted for the interested representation it is, once its historical contingency is fully acknowledged, it is no longer necessary to mask it. Freed from this burden of…dissimulation…the map will be able to assume its truest character, that of instrument for…data processing, that of instrument for…reasoning about quantitative information, that of instrument for…persuasive argument (Wood, 1992, p. 182)

If this is the case for geographic maps, then perhaps it is even more so for concept maps. Over the past several decades as I developed and worked with the idea of concept mapping described here, I frequently heard from my colleagues questions about the "validity" of the maps that were being generated. As a methodologist, I too was interested in this question and in several instances (Dumont, 1989; Trochim, 1993) attempted to address it, with varying degrees of success. It has not been clear how, exactly, one might methodologically best determine the accuracy or validity of these maps.

This brief sojourn into the land of the cartographer suggests that the validity question may be misplaced. It argues that every map is inherently inaccurate, and perhaps even worse (or is it better?) that the issue of accuracy is not even a very interesting or exciting one. In some sense, this recognition doesn't diminish the power of maps, geographic or otherwise. Perhaps instead it frees the maps up, in Wood's language, to do what they do best. The map is a powerful suggestive device. No one map can ever carry the weight of accuracy. Multiple maps based on multiple and diverse biases, perspectives and interests, aren't diminished because they are different. Potentially it is those differences that add value, that deepen insight, that increase our knowledge of the world. They assume their "… truest character, that of instrument for…data processing, that of instrument for…reasoning about quantitative information, that of instrument for…persuasive argument."

References

Anderberg, M.R. (1973). Cluster analysis for applications. New York, NY: Academic Press.

Bragg, L.R. and Grayson, T.E. (1993). Reaching consensus on outcomes: Lessons learned about concept mapping. Paper presented at the Annual Conference of the American Evaluation Association, Dallas, TX.

Buzan, T. with Buzan, B. (1993). The Mindmap Book: Radiant Thinking, The Major Evolution in Human Thought. BBC Books, London.

Caracelli, V. (1989). Structured conceptualization: A framework for interpreting evaluation results. Evaluation and Program Planning. 12, 1, 45-52.

Cook, J. (1992). Modeling staff perceptions of a mibile job support program for persons with severe mental illness. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Cooksy, L. (1989). In the eye of the beholder: Relational and hierarchical structures in conceptualization. Evaluation and Program Planning. 12, 1, 59-66.

Davis, J. (1989). Construct validity in measurement: A pattern matching approach. Evaluation and Program Planning. 12, 1, 31-36.

Davison, M.L. (1983). Multidimensional scaling. New York, John Wiley and Sons.

Dumont, J. (1989). Validity of multidimensional scaling in the context of structured conceptualization. Evaluation and Program Planning. 12, 1, 81-86.

Everitt, B. (1980). Cluster Analysis. 2nd Edition, New York, NY: Halsted Press, A Division of John Wiley and Sons.

Galvin, P.F. (1989). Concept mapping for planning and evaluation of a Big Brother/Big Sister program. Evaluation and Program Planning. 12, 1, 53-58.

Grayson, T.E. (1992). Practical issues in implementing and utilizing concept mapping. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Grayson, T.E. (1993). Empowering key stakeholders in the strategic planning and development of an alternative school program for youth at risk of school behavior. Paper presented at the Annual Conference of the American Evaluation Association, Dallas, TX.

Gurowitz, W.D., Trochim, W. and Kramer, H. (1988). A process for planning. The Journal of the National Association of Student Personnel Administrators, 25, 4, 226-235.

Harley, J.B. (1990). Text and contexts in the interpretation of early maps. In david Buisseret (Ed.) From Sea Charts to Satellite Images: Interpreting North American History Through Maps. University of Chicago Press. Chicago.

Imhoff, E. (1982). Cartographic Relief Representation. de gruyter Press, Berlin.

Kane, T.J. (1992). Using concept mapping to identify provider and consumer issues regarding housing for persons with severe mental illness. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Keith, D. (1989). Refining concept maps: Methodological issues and an example. Evaluation and Program Planning. 12, 1, 75-80.

Kohler, P.D. (1992). Services to students with disabilities in postsecondary education settings: Identifying program outcomes. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Kohler, P.D. (1993). Serving students with disabilities in postsecondary education settings: Using program outcomes for planning, evaluation and empowerment.Paper presented at the Annual Conference of the American Evaluation Association, Dallas, TX.

Kruskal, J.B. and Wish, M. (1978). Multidimensional Scaling. Beverly Hills, CA: Sage Publications.

Lassegard, E. (1992). Assessing the reliability of the concept mapping process. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Lassegard, E. (1993). Conceptualization of consumer needs for mental health services.Paper presented at the Annual Conference of the American Evaluation Association, Dallas, TX.

Linton, R. (1989). Conceptualizing feminism: Clarifying social science concepts. Evaluation and Program Planning. 12, 1, 25-30.

Mannes, M. (1989). Using concept mapping for planning the implementation of a social technology. Evaluation and Program Planning. 12, 1, 67-74.

Marquart, J.M. (1988). A pattern matching approach to link program theory and evaluation data: The case of employer-sponsored child care. Unpublished doctoral dissertation, Cornell University, Ithaca, New York.

Marquart, J.M. (1989). A pattern matching approach to assess the construct validity of an evaluation instrument. Evaluation and Program Planning. 12, 1, 37-44.

Marquart, J.M. (1992). Developing quality in mental health services: Perspectives of administrators, clinicians, and consumers. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Marquart, J.M., Pollak, L. and Bickman, L. (1993). Quality in intake assessment and case management: Perspectives of administrators, clinicians and consumers. In R. Friedman et al. (Eds.), A system of care for children's mental health: Organizing the research base. Tampa: Florida Mental Health Institute, University of South Florida.

McLinden, D. J. & Trochim, W.M.K. (In Press). From Puzzles to Problems: Assessing the Impact of Education in a Business Context with Concept Mapping and Pattern Matching. In J. Phillips (Ed.), Return on investment in human resource development: Cases on the economic benefits of HRD - Volume 2. Alexandria, VA: American Society for Training and Devlopment.

Mead, J.P. and Bowers, T.J. (1992). Using concept mapping in formative evaluations. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Mercer, M.L. (1992). Brainstorming issues in the concept mapping process. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Novak, J.D. (1993). How do we learn our lesson? Taking students through the process. The Science Teacher, 60, 3, 50-55.

Novak, J.D. and Gowin, D.B. (1985). Learning How To Learn. Cambridge, Cambridge University Press.

Nunnally, J.C. (1978). Psychometric Theory. (2nd. Ed.). New York, McGraw Hill.

Osborn, A.F. (1948). Your Creative Power. New York, NY: Charles Scribner.

Penney, N.E. (1992). Mapping the conceptual domain of provider and consumer expectations of inpatient mental health treatment: New York Results. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Romney, A.K., Weller, S.C. and Batchelder, W.H. (1986). Culture as consensus: A theory of culture and informant accuracy. American Anthropologist, 88, 2, 313-338.

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

Ryan, L. and Pursley, L. (1992). Using concept mapping to compare organizational visions of multiple stakeholders. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

SenGupta, S. (1993). A mixed-method design for practical purposes: Combination of questionnaire(s), interviews, and concept mapping.Paper presented at the Annual Conference of the American Evaluation Association, Dallas, TX.

Shern, D.L. (1992). Documenting the adaptation of rehabilitation technology to a core urban, homeless population with psychiatric disabilities: A concept mapping approach. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Shern, D.L., Trochim, W. and LaComb, C.A. (1995). The use of concept mapping for assessing fidelity of model transfer: An example from psychiatric rehabilitation. Evaluation and Program Planning, 18, 2.

Trochim, W. (1985). Pattern matching, validity, and conceptualization in program evaluation. Evaluation Review, 9, 5, 575-604.

Trochim, W. (Ed.) (1989). A Special Issue of Evaluation and Program Planning on Concept Mapping for Planning and Evaluation, 12.

Trochim, W. (1989a). An introduction to concept mapping for planning and evaluation. Evaluation and Program Planning, 12, 1, 1-16.

Trochim, W. (1989b). Concept mapping: Soft science or hard art? Evaluation and Program Planning, 12, 1, 87-110.

Trochim, W. (1989c). Outcome pattern matching and program theory. Evaluation and Program Planning, 12, 4, 355-366.

Trochim, W. (1990). Pattern matching and program theory. In H.C. Chen (Ed.), Theory-Driven Evaluation. New Directions for Program Evaluation, San Francisco, CA: Jossey-Bass.

Trochim, W. and Cook, J. (1992). Pattern matching in theory-driven evaluation: A field example from psychiatric rehabilitation. in H. Chen and P.H. Rossi (Eds.) Using Theory to Improve Program and Policy Evaluations. Greenwood Press, New York, 49-69.

Trochim, W. (1993). Reliability of Concept Mapping. Paper presented at the Annual Conference of the American Evaluation Association, Dallas, Texas, November.

Trochim, W. (1996). Criteria for evaluating graduate programs in evaluation. Evaluation News and Comment: The Magazine of the Australasian Evaluation Society, 5, 2, 54-57.

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

Trochim, W., Cook, J. and Setze, R. (1994). Using concept mapping to develop a conceptual framework of staff's views of a supported employment program for persons with severe mental illness. Consulting and Clinical Psychology, 62, 4, 766-775.

Valentine, K. (1989). Contributions to the theory of care. Evaluation and Program Planning. 12, 1, 17-24.

Valentine, K. (1992). Mapping the conceptual domain of provider and consumer expectations of inpatient mental health treatment: Wisconsin results. Paper presented at the Annual Conference of the American Evaluation Association, Seattle, WA.

Weller S.C. and Romney, A.K. (1988). Systematic Data Collection. Newbury Park, CA, Sage Publications.

Wilford, J.N. (1982). The Mapmakers: the Story of the great Pioneers in Cartography from Antiquity to the Space Age. New York, Vintage Books.

Wood, D. (1992) . The Power of Maps. The Guilford Press. New York.

Table 1. Brainstormed Statements

1. Increase membership of program staff and non-researchers.
2. Develop a logo.
3. Include public-trust businesses in organization, such as public utilities or waste management, which use GIF.
4. Recruit people working on the fringe of evaluation, e.g., people in assessment, training, technical assistance, etc.
5. Have skill development sessions more than once every other month.
6. Increase the diversity of membership.
7. Involve other local governments (counties, city, metro).
8. Meet with Board of Directors of agencies to promote evaluation.
9. Have annual conferences.
10. Have meetings during the day to accommodate the working researcher with family responsibilities.
11. Have committee to produce "press releases" about research in the area.
12. Provide a variety of for-fee events.
13. Involve evaluators in other parts of the state.
14. Provide information about available grants.
15. Create linkage with our neighbors to the north, and see what is going on regionally.
16. Have more facilitated roundtables.
17. Reach out to organizations receiving federal funds where evaluation is required.
18. Have follow-up information in newsletter about research opportunities originally announced via OPEN.
19. Provide feedback on the annual AEA meeting for those of us not able to attend.
20. Include "panels" as a program format.
21. Organize mini-conference.
22. Have no-host social hour at a variety of locations.
23. Enhance linkage with other research organizations.
24. Develop a network to include programs.
25. Develop linkages with police department to promote evaluation.
26. Work on ways to include evaluators from other parts of the state.
27. Enhance linkage with American Evaluation Association.
28. Include a brief bio on each OPEN member.
29. Hold events outside the downtown Portland area.
30. Develop a newsletter, email or otherwise.
31. Provide information about available contracts.
32. Continue the bi-monthly presentations
33. Do outreach to promote evaluation in the public and private sectors.
34. Develop special interest subgroups within the membership.
35. Use research knowledge for policy change.
36. Do more outreach.
37. Include interactive discussion with audience at program sessions.
38. Join state and county advisory boards as representative of OPEN.
39. Increase membership of auditors and accountants.
40. Continue email announcements of employment opportunity.
41. Distribute the results of research to members and agencies.
42. Circulate a list of members who are expert in various techniques and who are willing to do brief consulting with others.
43. Have "special interest" or "special topic" brown bags.
44. Hear more from local evaluators about what their workload is like, including information about on-going projects.
45. Have a "best practices" model available at conferences or workshops.
46. Develop student scholarship fund for conferences.
47. Develop a pool of volunteers to help agencies needing evaluation but with out funding.

48. Provide a summary of the skill-building sessions in the newsletter.
49. Develop a better communication system for members and nonmembers.
50. Develop student scholarship fund for seed money for research (e.g., dissertations).
51. Continue to offer information/knowledge development sessions.
52. Add linkage with national think tanks.
53. Provide community service by offering evaluation to agencies with no funding for such activities.
54. Develop a speakers committee.
55. Do a better job of advertising the benefits of membership in OPEN.
56. Include students and other disenfranchised groups by offering free membership.
57. Have an annual picnic for members and nonmembers.
58. Increase membership and participation of people of color.
59. Coordinate communication between email, webpage and newsletter.
60. Broaden scope of organization to include all research, not just program evaluation.
61. Increase diversity of membership to include business sectors.
62. Develop a mentor program for students and recent graduates.
63. Have a student division.
64. Develop a mentoring program for recent graduates.
65. Use "lecture" format for the program topics.
66. Develop way to monitor social health in the tri-county area and sell idea to Commissioners to collect data annually.
67. Develop technical assistance program for members to help other members.
68. Arrange better access to libraries for OPEN members.
69. Work to develop collaboration between OPEN and agencies needing research and evaluation.
70. Create an interactive chat-room on the website.
71. Develop scholarship for internships.
72. Obtain non-profit status.
73. Work to foster collaboration between members.
74. Plan and schedule the bi-monthly meetings an entire year in advance.
75. Get the membership more active and participatory.
76. Have meetings someplace other than the Lucky Lab.
77. Include "debate" as a format of the program.
78. Provide student scholarships to AEA conference.
79. Arrange for Continuing Education Credit for the programs.
80. Membership directory listings include areas in which members would like to develop their skills.
81. Double the size of the membership.
82. Provide a variety of no-charge events.
83. Do outreach to higher education and public schools.

Table 2. Demographics

Name: Employment Where is your primary place of employment:	Please select one: government education non-profit agency research consultant student other
Name: Degree What is the highest academic degree earned?	Please select one: BA/BS degree Master's degree Ph.D./Doctorate Other
Name: Discipline What is your primary discipline?	Please select one: Agriculture Business Economy Education Evaluation Health Human Services Political Sciences Psychology Sociology Social Work Other
Name: Years in Research How many years of experience have you had in research or related activities?	Enter a value between 0 and 75:
Name: OPEN Member Are you a member of OPEN?	Please select one: Yes No
Name: AEA Member Are you a member of the American Evaluation Association?	Please select one: Yes No
Name: Gender What is your gender?	Please select one: Female Male
Name: Residence In what city do you reside?	Please select one: Portland Portland east suburb Portland west suburb Corvallis Salem Eugene Oregon - Other Washington State Other

Table 3. Statements by Cluster in Descending Order by Average Rating

Recruitment

58) Increase membership and participation of people of color.	3.83
6) Increase the diversity of membership.	3.78
13) Involve evaluators in other parts of the state.	3.72
75) Get the membership more active and participatory.	3.61
55) Do a better job of advertising the benefits of membership in OPEN.	3.17
81) Double the size of the membership.	3.11
60) Broaden scope of organization to include all research, not just program evaluation.	2.94
4) Recruit people working on the fringe of evaluation, e.g., people in assessment, training, technical assistance, etc.	2.78
63) Have a student division.	2.78
1) Increase membership of program staff and non-researchers.	2.67
56) Include students and other disenfranchised groups by offering free membership.	2.67
61) Increase diversity of membership to include business sectors.	2.61
39) Increase membership of auditors and accountants.	2.22
Average Rating:	3.07

Linkages

7) Involve other local governments (counties, city, metro).	3.61
26) Work on ways to include evaluators from other parts of the state.	3.61
15) Create linkage with our neighbors to the north, and see what is going on regionally.	3.56
24) Develop a network to include programs.	2.72
3) Include public-trust businesses in organization, such as public utilities or waste management, which use GIF.	2.41
Average Rating:	3.18

PR & Community Service

35) Use research knowledge for policy change.	3.22
47) Develop a pool of volunteers to help agencies needing evaluation but with out funding.	3.11
53) Provide community service by offering evaluation to agencies with no funding for such activities.	3.00
66) Develop way to monitor social health in the tri-county area and sell idea to Commissioners to collect data annually.	2.89
54) Develop a speakers committee.	2.83
11) Have committee to produce "press releases" about research in the area.	2.72
2) Develop a logo.	2.17
Average Rating:	2.85

Outreach

69) Work to develop collaboration between OPEN and agencies needing research and evaluation.	3.89

33) Do outreach to promote evaluation in the public and private sectors.	3.44
23) Enhance linkage with other research organizations.	3.33
27) Enhance linkage with American Evaluation Association.	3.33
17) Reach out to organizations receiving federal funds where evaluation is required.	3.28
36) Do more outreach.	3.28
52) Add linkage with national think tanks.	3.28
83) Do outreach to higher education and public schools.	3.11
8) Meet with Board of Directors of agencies to promote evaluation.	3.06
38) Join state and county advisory boards as representative of OPEN.	2.61
25) Develop linkages with police department to promote evaluation.	2.56
Average Rating:	3.20

Programs & Events

51) Continue to offer information/knowledge development sessions.	4.56
32) Continue the bi-monthly presentations	4.06
43) Have "special interest" or "special topic" brown bags.	3.89
9) Have annual conferences.	3.83
10) Have meetings during the day to accommodate the working researcher with family responsibilities.	3.44
21) Organize mini-conference.	3.44
45) Have a "best practices" model available at conferences or workshops.	3.41
37) Include interactive discussion with audience at program sessions.	3.22
44) Hear more from local evaluators about what their workload is like, including information about on-going projects.	3.22
72) Obtain non-profit status.	3.17
20) Include "panels" as a program format.	3.11
16) Have more facilitated roundtables.	3.06
77) Include "debate" as a format of the program.	3.00
76) Have meetings someplace other than the Lucky Lab.	2.94
82) Provide a variety of no-charge events.	2.89
29) Hold events outside the downtown Portland area.	2.83
74) Plan and schedule the bi-monthly meetings an entire year in advance.	2.78
65) Use "lecture" format for the program topics.	2.67
12) Provide a variety of for-fee events.	2.61
22) Have no-host social hour at a variety of locations.	2.61
5) Have skill development sessions more than once every other month.	2.56
57) Have an annual picnic for members and nonmembers.	2.22
Average Rating:	3.16

Communication

40) Continue email announcements of employment opportunity.	4.56
48) Provide a summary of the skill-building sessions in the newsletter.	3.83
31) Provide information about available contracts.	3.78
14) Provide information about available grants.	3.67
19) Provide feedback on the annual AEA meeting for those of us not able to attend.	3.61
30) Develop a newsletter, email or otherwise.	3.44

59) Coordinate communication between email, webpage and newsletter.	3.11
18) Have follow-up information in newsletter about research opportunities originally announced via OPEN.	2.61
70) Create an interactive chat-room on the website.	2.00
Average Rating:	3.40

Services

42) Circulate a list of members who are expert in various techniques and who are willing to do brief consulting with others.	3.89
73) Work to foster collaboration between members.	3.78
41) Distribute the results of research to members and agencies.	3.72
67) Develop technical assistance program for members to help other members.	3.72
34) Develop special interest subgroups within the membership.	3.44
28) Include a brief bio on each OPEN member.	3.06
49) Develop a better communication system for members and nonmembers.	3.06
68) Arrange better access to libraries for OPEN members.	2.94
79) Arrange for Continuing Education Credit for the programs.	2.89
80) Membership directory listings include areas in which members would like to develop their skills.	2.56
Average Rating:	3.31

Mentoring/Scholarships

64) Develop a mentoring program for recent graduates.	3.22
62) Develop a mentor program for students and recent graduates.	3.11
46) Develop student scholarship fund for conferences.	2.94
71) Develop scholarship for internships.	2.39
50) Develop student scholarship fund for seed money for research (e.g., dissertations).	2.28
78) Provide student scholarships to AEA conference.	2.28
Average Rating:	2.70

Table 4. Statements in Descending Order by Average Rating

40)	Continue email announcements of employment opportunity.	4.56
51)	Continue to offer information/knowledge development sessions.	4.56
32)	Continue the bi-monthly presentations	4.06
42)	Circulate a list of members who are expert in various techniques and who are willing to do brief consulting with others.	3.89
43)	Have "special interest" or "special topic" brown bags.	3.89
69)	Work to develop collaboration between OPEN and agencies needing research and evaluation.	3.89
9)	Have annual conferences.	3.83
48)	Provide a summary of the skill-building sessions in the newsletter.	3.83
58)	Increase membership and participation of people of color.	3.83
6)	Increase the diversity of membership.	3.78
31)	Provide information about available contracts.	3.78
73)	Work to foster collaboration between members.	3.78
13)	Involve evaluators in other parts of the state.	3.72
41)	Distribute the results of research to members and agencies.	3.72
67)	Develop technical assistance program for members to help other members.	3.72
14)	Provide information about available grants.	3.67
7)	Involve other local governments (counties, city, metro).	3.61
19)	Provide feedback on the annual AEA meeting for those of us not able to attend.	3.61
26)	Work on ways to include evaluators from other parts of the state.	3.61
75)	Get the membership more active and participatory.	3.61
15)	Create linkage with our neighbors to the north, and see what is going on regionally.	3.56
10)	Have meetings during the day to accommodate the working researcher with family responsibilities.	3.44
21)	Organize mini-conference.	3.44
30)	Develop a newsletter, email or otherwise.	3.44
33)	Do outreach to promote evaluation in the public and private sectors.	3.44
34)	Develop special interest subgroups within the membership.	3.44
45)	Have a "best practices" model available at conferences or workshops.	3.41
23)	Enhance linkage with other research organizations.	3.33
27)	Enhance linkage with American Evaluation Association.	3.33
17)	Reach out to organizations receiving federal funds where evaluation is required.	3.28
36)	Do more outreach.	3.28
52)	Add linkage with national think tanks.	3.28
35)	Use research knowledge for policy change.	3.22
37)	Include interactive discussion with audience at program sessions.	3.22
44)	Hear more from local evaluators about what their workload is like, including information about on-going projects.	3.22
64)	Develop a mentoring program for recent graduates.	3.22
55)	Do a better job of advertising the benefits of membership in OPEN.	3.17
72)	Obtain non-profit status.	3.17
20)	Include "panels" as a program format.	3.11
47)	Develop a pool of volunteers to help agencies needing evaluation but with out funding.	3.11
59)	Coordinate communication between email, webpage and newsletter.	3.11
62)	Develop a mentor program for students and recent graduates.	3.11
81)	Double the size of the membership.	3.11
83)	Do outreach to higher education and public schools.	3.11
8)	Meet with Board of Directors of agencies to promote evaluation.	3.06
16)	Have more facilitated roundtables.	3.06
28)	Include a brief bio on each OPEN member.	3.06
49)	Develop a better communication system for members and nonmembers.	3.06

53)	Provide community service by offering evaluation to agencies with no funding for such activities.	3.00
77)	Include "debate" as a format of the program.	3.00
46)	Develop student scholarship fund for conferences.	2.94
60)	Broaden scope of organization to include all research, not just program evaluation.	2.94
68)	Arrange better access to libraries for OPEN members.	2.94
76)	Have meetings someplace other than the Lucky Lab.	2.94
66)	Develop way to monitor social health in the tri-county area and sell idea to Commissioners to collect data annually.	2.89
79)	Arrange for Continuing Education Credit for the programs.	2.89
82)	Provide a variety of no-charge events.	2.89
29)	Hold events outside the downtown Portland area.	2.83
54)	Develop a speakers committee.	2.83
4)	Recruit people working on the fringe of evaluation, e.g., people in assessment, training, technical assistance, etc.	2.78
63)	Have a student division.	2.78
74)	Plan and schedule the bi-monthly meetings an entire year in advance.	2.78
11)	Have committee to produce "press releases" about research in the area.	2.72
24)	Develop a network to include programs.	2.72
1)	Increase membership of program staff and non-researchers.	2.67
56)	Include students and other disenfranchised groups by offering free membership.	2.67
65)	Use "lecture" format for the program topics.	2.67
12)	Provide a variety of for-fee events.	2.61
18)	Have follow-up information in newsletter about research opportunities originally announced via OPEN.	2.61
22)	Have no-host social hour at a variety of locations.	2.61
38)	Join state and county advisory boards as representative of OPEN.	2.61
61)	Increase diversity of membership to include business sectors.	2.61
5)	Have skill development sessions more than once every other month.	2.56
25)	Develop linkages with police department to promote evaluation.	2.56
80)	Membership directory listings include areas in which members would like to develop their skills.	2.56
3)	Include public-trust businesses in organization, such as public utilities or waste management, which use GIF.	2.41
71)	Develop scholarship for internships.	2.39
50)	Develop student scholarship fund for seed money for research (e.g., dissertations).	2.28
78)	Provide student scholarships to AEA conference.	2.28
39)	Increase membership of auditors and accountants.	2.22
57)	Have an annual picnic for members and nonmembers.	2.22
2)	Develop a logo.	2.17
70)	Create an interactive chat-room on the website.	2.00