CONCEPT MAPPING FOR EVALUATION AND PLANNING

William M.K. Trochim
Guest Editor

A Special Issue of
Evaluation and Program Planning

PERGAMON PRESS
NEW YORK • OXFORD • BEIJING • FRANKFURT • SÃO PAULO • SYDNEY • TOKYO • TORONTO

Pergamon Press Offices:

U.S.A.	Pergamon Press, Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
U.K.	Pergamon Press, Headington Hill Hall, Oxford OX3 0BW, England
PEOPLE'S REPUBLIC OF CHINA	Pergamon Press, Qianmen Hotel, Beijing, People's Republic of China
FEDERAL REPUBLIC OF GERMANY	Pergamon Press, Hammerweg 6, D-6242 Kronberg, Federal Republic of Germany
BRAZIL	Pergamon Editora, Rua Eça de Queiros, 346, CEP 04011, São Paulo, Brazil
AUSTRALIA	Pergamon Press Australia, P.O. Box 544, Potts Point, N.S.W. 2011, Australia
JAPAN	Pergamon Press, 8th Floor, Matsuoka Central Building, 1-7-1 Nishishinjuku, Shinjuku-ku, Tokyo 160, Japan
CANADA	Pergamon Press Canada, Suite 271, 253 College Street, Toronto, Ontario, M5T 1R5, Canada

Published as Volume 12, Number 1, 1989 of *Evaluation and Program Planning* and also available to non-subscribers.

Printed in the United States of America

EVALUATION AND PROGRAM PLANNING

Volume 12, Number 1

1989

SPECIAL ISSUE:
CONCEPT MAPPING FOR EVALUATION AND PLANNING

William M.K. Trochim
Guest Editor

CONTENTS

(Continued on next page)

Evaluation and Program Planning is indexed in Current Contents/Social & Behavioral Sciences, Current Contents/Health Services Administration, Soc. Sci. Citation Index, Psychol. Abstr., PsycINFO, Sociol. Abstr., Adolescent Mental Health Abstr., ERIC/CIJE

ISSN 0149-7189
(593)

(contents continued)

Evaluation and Program Planning, Vol. 12, pp. 1-16, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

AN INTRODUCTION TO CONCEPT MAPPING FOR PLANNING AND EVALUATION

WILLIAM M.K. TROCHIM

Cornell University

ABSTRACT

Concept mapping is a type of structured conceptualization which can be used by groups to develop a conceptual framework which can guide evaluation or planning. In the typical case, six steps are involved: (1) Preparation *(including selection of participants and development of focus for the conceptualization); (2) the* Generation of statements*; (3) the* Structuring of statements*; (4) the* Representation of Statements *in the form of a concept map (using multidimensional scaling and cluster analysis); (5) the* Interpretation of maps*; and, (6) the* Utilization of maps. *Concept mapping: encourages the group to stay on task; results relatively quickly in an interpretable conceptual framework; expresses this framework entirely in the language of the participants; yields a graphic or pictorial product which simultaneously shows all major ideas and their interrelationships; and often improves group or organizational cohesiveness and morale. This paper describes each step in the process, considers major methodological issues and problems, and discusses computer programs which can be used to accomplish the process.*

Probably the most difficult step in a planning or evaluation project is the first one—everything which follows depends on how well the project is initially conceptualized. Conceptualization in this sense refers to the articulation of thoughts, ideas, or hunches and the representation of these in some objective form. In a planning process, we typically wish to conceptualize the major goals and objectives, needs, resources, and capabilities or other dimensions which eventually constitute the elements of a plan. In evaluation, we may want to conceptualize the programs or treatments, samples, settings, measures, and outcomes which we believe are relevant.

This special issue of *Evaluation and Program Planning* extends earlier work by Trochim and Linton (1986) who proposed a general framework for structured conceptualization and showed how specific conceptualization processes can be devised to assist groups in the theory and concept formation stages of planning and evaluation. The papers presented here focus on one specific type of structured conceptualization process which we term "concept mapping." In concept mapping, ideas are represented in the form of a picture or map. To construct the map, ideas first have to be described or generated, and the interrelationships between them articulated. Multivariate statistical techniques—multidimensional scaling and cluster analysis—are then applied to this information and the results are depicted in map form. The content of the map is entirely determined by the group. They brainstorm the initial ideas, provide information about how these ideas are related, interpret the results of the analyses, and decide how the map is to be utilized.

The process described here is not the only way to accomplish concept mapping. For instance, Novak and Gowin (1984) suggest that concept maps be drawn "free-hand" after an initial articulation of the major ideas and classification of them into hierarchical concepts. In a similar manner, Rico (1983) has advocated "free-hand" concept mapping or drawing as a useful method for developing a conceptual framework for writing. These and other approaches have value for planning and evaluation, but fall outside of the scope of this paper. The major differences between the method described here and other concept mapping processes are: this method is particularly appropriate for group use—the method generates a group aggregate map; it utilizes multivariate data analyses to construct the

Requests for reprints should be sent to William M.K. Trochim, Department of Human Service Studies, MVR Hall, Cornell University, Ithaca, NY 14853.

maps; and it generates interval-level maps which have some advantages for planning and evaluation, especially through pattern matching as described later. Despite these differences, this paper should be viewed as a clear call for the importance of further exploration of *any* processes which improve conceptualization in planning and evaluation. Throughout the papers in this volume, however, the term "concept mapping" should be understood to refer only to the process described here, and its variations.

Group concept mapping is consistent with the growing interest in the role of theory in planning and evaluation. In evaluation, for instance, this interest is evidenced in writings on the importance of program theory (Bickman, 1986; Chen & Rossi, 1983, 1987; Rossi & Chen, in press); in the increased emphasis on the importance of studying causal process (Mark, 1986); in the recognition of the central role of judgment—especially theory-based judgment—in research (Cordray, 1986; Einhorn & Hogarth, 1986); and, in the thinking of critical multiplism (Shadish, Cook, & Houts, 1986) which emphasizes the role of theory in selecting and guiding the analysis of multiple operationalizations. Concept mapping can be viewed as one way to articulate theory in these contexts. In planning, conceptualization has had somewhat more attention and is evidenced in the sometimes daunting proliferation of different planning models and methods of conceptualizing (Dunn, 1981).

This paper introduces the concept mapping process, suggests some of the major technical or methodological issues which are involved, and offers some suggestions about computer programs which can be used to accomplish concept mapping. The remaining papers provide numerous examples of the use of this process and its variations in a wide variety of contexts and considers in greater detail some of the methodological issues which face researchers in this area.

THE CONCEPT MAPPING PROCESS

The term "structured conceptualization" refers to *any* process which can be described as a sequence of concrete operationally-defined steps and which yields a conceptual representation (Trochim & Linton, 1986). The specific concept mapping process described here and discussed throughout this volume is considered only one of many possible structured conceptualization processes. This process can be used whenever there is a group of people who wish to develop a conceptual framework for evaluation or planning, where the framework is displayed in the form of a concept map. A concept map is a pictorial representation of the group's thinking which displays all of the ideas of the group relative to the topic at hand, shows how these ideas are related to each other and, optionally, shows which ideas are more relevant, important, or appropriate.

The scenario within which concept mapping is applied assumes that there is an identifiable group responsible for guiding the evaluation or planning effort. Depending on the situation, this group might consist of the administrators, staff or members of the board of an organization; community leaders or representatives of relevant constituency groups; academicians or members of the policy making community; funding agents or representatives of groups with oversight responsibility; representatives of relevant client populations; or combinations of these. The concept mapping process is guided by a facilitator who could be an outside consultant or an internal member of the group responsible for the planning or evaluation effort. The facilitator's role is only to manage the process—the content, interpretation and utilization of the concept map are determined entirely by the group.

An overview of the concept mapping process is provided in Figure 1. The figure shows six steps which are followed in developing a useful group concept map. Each of these steps will be discussed in some detail and illustrated with data from a concept mapping process which was conducted in York County, Maine, to assist representatives of a number of county human service agencies to develop a conceptual framework for planning services for the elderly. While the major focus of this project was on planning, some comments will be offered regarding the potential use of this conceptualization for evaluation purposes.

Step 1: Preparation

There are two major tasks which must be undertaken prior to commencement of the actual group process. First, the facilitator must work with the parties involved to decide on who will participate in the process. Second, the facilitator must then work with the participants or a subgroup to decide on the specific focus for the conceptualization.

Selecting the Participants. One of the most important tasks which the facilitator addresses is who will participate in the concept mapping process. Our experience has been that a conceptualization is best when it includes a wide variety of relevant people. If we are conducting strategic planning for a human service organization, we might include administrative staff, service staff, board members, clients, and relevant members of community groups. In a program evaluation context, we might similarly include administrators, program staff, clients, social science theorists, community members, and rele-

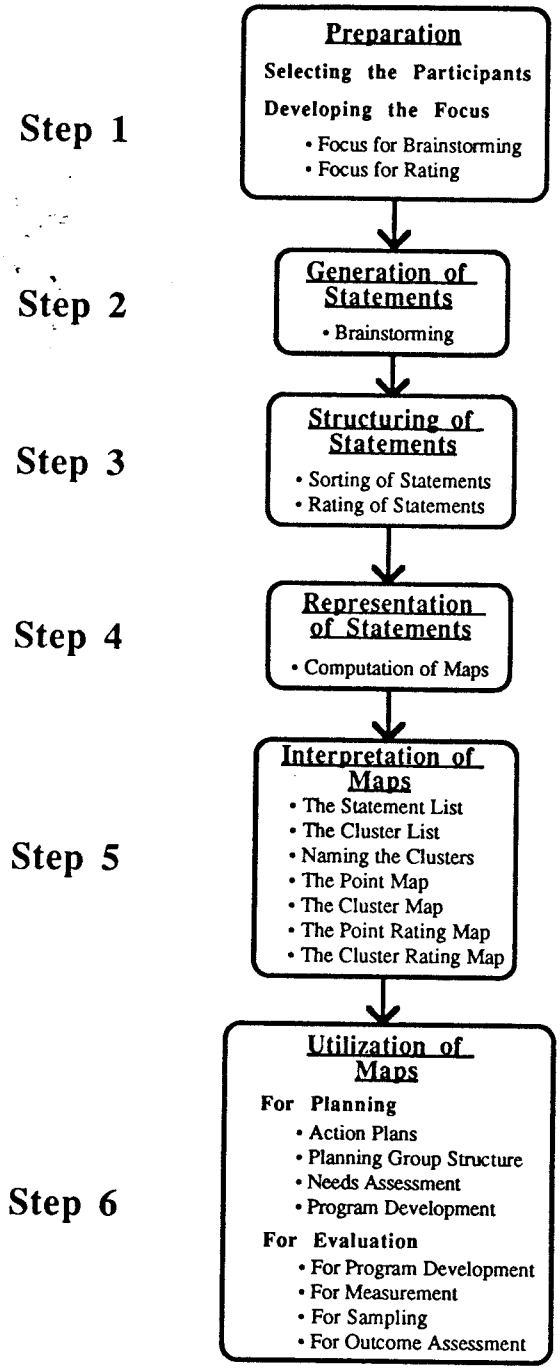


Figure 1. The concept mapping process.

vant funding agent representatives. Broad heterogeneous participation helps to insure that a wide variety of viewpoints will be considered and encourages a broader range of people to "buy into" the conceptual framework which results.

In some situations, however, we have used relatively small homogenous groups for the conceptualization process. For instance, if an organization is beginning a strategic planning effort and would like to lay out quickly some of the major concepts around which the planning will be based, they might prefer to use a relatively small group of administrators and organizational staff members. The obvious advantage of doing this is that it is logistically simpler to get people together for meetings if they are all on the staff of the organization. This type of group works well when a quick conceptualization framework is desired, but in general we would recommend a broader sampling of opinion.

In some contexts it might be reasonable to use some random sampling scheme to select participants from a larger defined population. This is most useful when one wishes to argue that the resulting concept map is generalizable to some larger population of interest. Simple random sampling schemes, of course, run the risk of underrepresenting minority groups from the population and so, if sampling is used, it will typically be best to attempt either some form of stratified random sampling or purposive sampling for heterogeneity.

There is no strict limit on the number of people who can be involved in concept mapping. It is feasible, with some process modifications, for an individual to conduct a conceptualization alone (see Dumont, this volume, for an example). At the other extreme, we have worked with groups as large as 75–80 people in the process. Typically, we have had between 10 and 20 people in most of our studies and this seems to be a workable number. Groups of that size insure a variety of opinions and still enable good group discussion and interpretation.

It is also not necessary that all participants take part in every step of the process. One might, for instance, have a relatively small group do the generation (e.g. brainstorming) step, a much larger group perform the structuring (i.e., sorting and rating) and a small group for interpretation and utilization. In general, however we have found that concept maps are better understood by people who have participated in all phases of the process than by those who have only taken part in one or two steps.

In the York County study, the purpose of the conceptualization was to bring together a small group representatives from a number of local agencies which provide services to the elderly in order to develop framework for planning. There was also a strong interest in piloting the concept mapping process with the idea that it might possibly be applied later with a broader constituency group which included elderly persons from the county. Between 10 and 15 people participated in the two meetings including representatives of the United Way and several health and mental health organizations.

Developing the Focus. The second major preparation step involves developing the focus or domain of

conceptualization. There are two separate products which are desired here. First, the participants must define the focus for the brainstorming session. Second, the focus for ratings which are performed during the structuring step of the process needs to be developed. This essentially involves defining the dimension(s) on which each of the brainstormed statements will be rated.

It is essential that the focus for both the brainstorming and the ratings be worded as statements which give the specific instruction intended so that all of the participants can agree in advance. In developing both the brainstorming and rating focus statements, the facilitator usually has a meeting with the participants or some representative subgroup. In this meeting the facilitator discusses various alternatives for wording each focus and attempts to achieve a group consensus on final choices. For example, the brainstorming focus in a strategic planning process might be worded: "Generate short phrases or sentences which describe specific services which your organization might provide." Similarly, a rating focus for a program evaluation might be worded: "Rate each potential outcome on a seven point scale in terms of how strongly you think it will be affected by the program, where '1' means 'Not at all affected', '4' means 'Moderately affected' and '7' means 'Extremely affected'." The group should agree on the specific wording for each of these focus statements.

The Brainstorming Focus. For any brainstorming session, there are a variety of ways in which the focus can be stated. For instance, if we are interested in strategic planning, participants might focus on the goals of the organization, the mission of the organization, or the activities or services which the organization might provide. Similarly, in program evaluations they might focus on the nature of the program, the outcomes they would like to measure, or the types of people they would like to participate in the evaluation.

In defining the brainstorming focus, it is important to try to anticipate the types of statements which will result. For instance, one usually wants to avoid giving double-barreled focus statements like: "Generate short statements or sentences which describe the goals of our organization and the needs of our clients" because, when sorting, participants are likely to perceive these two categories as particularly distinct and consequently, they are likely to show up as two major clusters on the final concept map and obscure some of the finer relationships which might be of interest (see Keith, this volume, for a discussion of this issue). If both emphases are desired, it would probably be better to conduct two separate conceptualizations to address each issue. In general, it is best if the final set of brainstormed statements are "of a kind," that is, share the same level of conceptual generality and grammatic structure. For example, a brainstormed set which consists entirely of statements which describe specific services which an organization might provide and which are all worded in a similar manner (e.g., some are not in question form while others are worded as statements) would be best if the interest is in conceptualizing a taxonomy of potential services.

In the York County study, the brainstorming focus statement was a broad one: "Generate statements which describe the issues, problems, concerns or needs which the elderly have in York County." Although this may seem to violate the advice above on avoiding double-barreled statements, the participants felt that "issues, problems, concerns or needs" were essentially indistinguishable and that the brainstormed statements which resulted would constitute an interpretable, relatively homogeneous set.

The Rating Focus. In developing the focus for the ratings, one needs to consider what kind of information will be most useful. In planning, it is often useful to ask the participants to rate how important each brainstormed item is, or how much emphasis should be placed upon it in the planning process. In evaluation, it might be useful to ask them to rate how much effort should be given to various program components, or how much they believe each outcome is likely to be affected by the program.

In the York County study, the rating focus statement was: "Rate each statement on a 1 to 5 scale for how much priority it should be given in the planning process, where '1' equals the lowest priority and '5' equals the highest priority."

Step 2: Generation of Statements

Once the participants and focus statements have been defined, the actual concept mapping process begins with the generation of a set of statements which ideally should represent the entire conceptual domain for the topic of interest. In the typical case, brainstorming is used and the focus statement constitutes the prompt for the brainstorming session. The usual rules for brainstorming apply (Dunn, 1981; Osborn, 1948). That is, people are encouraged to generate lots of statements and are told that there should be no criticism or discussion regarding the legitimacy of statements which are generated during the session. Participants are encouraged to ask for clarification of any unfamiliar terms or jargon so that all who participate may understand what was intended by a given statement. Usually, the facilitator records the statements as they are generated so that all members of the group can see the set of statements as they evolve. We have done this by writing the statements on a blackboard, on sheets of newsprint, or by directly entering them into a computer program which is displayed on a large screen so that everyone

can see the statements. If the facilitator believes that there may be some participants who would be reluctant to state publicly some idea because of its controversial or potentially embarrassing nature, it may be desirable to also allow each participant to submit several statements anonymously on paper so that confidentiality will be preserved.

Theoretically, there is no limit to the number of statements which can be generated. However, large numbers of statements impose serious practical constraints. Based on our experience, we now limit the number of statements to one hundred or less. If the brainstorming session generates more than a hundred statements, we reduce the set. There are a number of ways in which this can be accomplished. The group as a whole or some subgroup can examine the set of statements for redundancies or ones which can be chosen to represent a set of others. In some instances, we have taken a simple random sample of statements from the larger set and had participants examine the selected set to be sure that no key ideas were being omitted. For instance, Linton (this volume) randomly selected 150 statements out of a larger set of 710 statements that were generated. On several occasions we have tried more formal thematic analysis of the text statements using a "key words in context" approach (Krippendorf, 1980; Stone, Dunphy, Smith, & Ogilvie, 1966) which seems promising.

Once a final set of statements has been generated, it is valuable for the group to examine the statements for editing considerations. Sometimes the wording of statements generated in a brainstorming session is awkward or technical jargon is not clear. In general, each statement should be consistent with what was called for in the brainstorming prompt and should be detailed enough so that every member of the group can understand the essential meaning of the statement.

There are many other ways to generate the conceptual domain than brainstorming. Sometimes a set of statements can be abstracted from existing text documents such as annual reports, internal organizational memos, interviews or field notes. For instance, Dumont (this volume) utilized the "documentary coding method" described by Wrightson (1976) to abstract statements or entities from interview records. In some cases there was no need to generate statements at all because the nature of the conceptualization dictates the elements of the conceptual domain. Thus, if the goal is to conceptualize the interrelationships between the set of 10 departments in an organization, we might simply use the 10 department names as the set of statements. Marquart (this volume) used a set of concepts derived from the literature on employer-sponsored child care as the set of statements for mapping. Caracelli (this volume) used the 100 items from the California Q-sort as the set of statements. We have also done some preliminary work to examine the use of outlines for generation where each entry in the outline would be considered a separate statement. One potential advantage of doing this is that we may be able to use the implicit structure of the outline headings and subheadings to structure the conceptual domain directly without asking people to sort the statements. Cooksy (this volume) describes some preliminary attempts to develop models which take outlines as input and estimate the similarity between statements (based on outline structure) as output.

In the York County study, a set of 95 statements was generated in the brainstorming session and are shown in Table 1. The statements describe a broad range of issues which were thought by participants to be salient to the elderly in York County.

Step 3: Structuring of Statements

Once we have a set of statements which describes the conceptual domain for a given focus, we minimally need to provide information about how the statements are related to each other. In addition, we often want to rate each statement on some dimension which is defined by the rating focus statement. Both of these tasks constitute the structuring of the conceptual domain.

Typically, we obtain information about interrelationships using an unstructured card sorting procedure (Rosenberg & Kim, 1975). Each of the brainstormed statements is printed on a separate 3 × 5 index card and the complete set of cards is given to each participant. Each person is then instructed to sort the cards into piles "in a way that makes sense to you." There are several restrictions placed on this procedure: each statement can only be placed in one pile (i.e., an item cannot be placed in two piles simultaneously); all statements cannot be put into a single pile; and, all statements cannot be put into their own pile (although some items may be sorted by themselves). Except for these conditions, people may pile the cards in any way that makes sense to them. Often the participants perceive that there may be several different ways to sort the cards, all of which make sense. To address this, we have either instructed participants to select the most sensible arrangement or, in some studies, have had each participant sort the cards several times.

When each person has completed the sorting task, the results must be combined across people. This is accomplished in two steps. First, the results of the sort for each person are put into a square table or matrix which has as many rows and columns as there are statements. All of the values of this matrix are either zero or one. A '1' indicates that the statements for that row and column were placed by that person together in a pile while a '0' indicates that they were not. This is illustrated in Figure 2 for a hypothetical person who sorted ten statements into 4 piles.

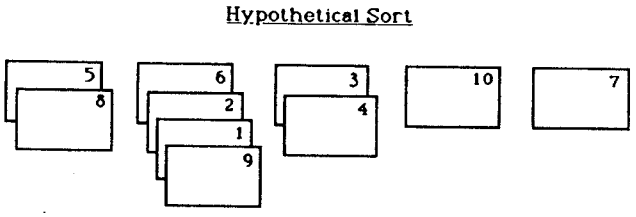
We can see in the figure that statements 5 and 8 were

TABLE 1
BRAINSTORMED STATEMENTS FROM YORK COUNTY ELDERLY PROJECT

1. lack of public transportation
2. lack of money
3. high cost of medication
4. high cost of health care
5. stigmatization of elderly persons
6. isolation
7. dependency—interpersonal
8. loss of close support
9. deterioration of mental capacity
10. physical disabilities
11. late life role changes
12. depression
13. inadequate home care services
14. lack of nursing homes
15. decreased functional status
16. loss of recreation
17. concern for mental wellness
18. emergency supervised housing
19. lack of socialization
20. fears about not being able to protect themselves
21. inability to cope with change
22. effects of the weather on the elderly
23. death of spouse
24. lack of funding for community care
25. lack of expertise to treat mentally ill elderly
26. confusing social service system
27. exploitation
28. physical abuse and neglect by caretakers
29. duplication of services
30. lack of outreach
31. losing your driver's license
32. Medicare limitations
33. nutrition services
34. expenses of maintaining family contacts
35. need for political advocacy
36. housing costs
37. different housing alternatives
38. over self-medication
39. over-medication by physicians
40. inability to understand or read medication instructions
41. need for independence
42. lack of awareness in mixing drugs and alcohol
43. expanding (exploding) elderly population
44. continuation of vocational satisfaction
45. lack of preparation for aging (including retirement)
46. issues around self-worth and self-esteem
47. changing role in the community
48. deterioration of family life
49. need for respite for caregivers
50. awareness of community resources
51. long-term care insurance
52. Medicare coverage being gradually "squeezed down"
53. lack of state and federal funding commitment to elderly
54. fear of prolonged dying
55. concern over being a burden
56. feeling of uselessness
57. feeling of being excluded by other generations or groups
58. desire for access to educational/recreational program
59. desire to be treated like everybody else
60. assistance with managing finances
61. job discrimination
62. lack of education on life changes
63. exclusion of elderly as an educational resource
64. desire to pass on traditions and knowledge/experiences
65. how do some elderly function so well? (exceptional elderly)
66. premature categorization as "old" or elderly
67. care for elderly persons with dementia
68. confusion between aging and illness
69. improved use of educational opportunities
70. access issues for disabled elderly
71. accessible education through creative programming
72. more diverse social opportunities
73. placement outside of home
74. means to maintain independence
75. in-home visits by physicians
76. clergy that are aware of community resources
77. improved public awareness and support of elderly needs
78. clergy who can make home visits
79. understanding and respect for sexuality
80. fear of being crazy
81. suicide among the elderly
82. alcoholism among the elderly
83. infantilization of elderly by providers
84. an institutional end (decline and death)
85. giving elderly persons the right to choose to die
86. fear of being abandoned
87. loss of legal control
88. loss of financial control
89. better understanding of elderly by legal profession
90. distress about the family issues of their children
91. high cost of home nursing care
92. role models for the elderly
93. depletion of resources by children
94. lack of support for family caregivers
95. law enforcement intervention when elderly refuse to be hospitalized.

sorted together in a pile. Therefore, in the table the row 5—column 8 and row 8—column 5 entries are '1'. Because statement 5 was not sorted with statement 6, the row 5—column 6 and row 6—column 5 entries are '0'. This individual matrix is termed a binary symmetric similarity matrix. Notice that all of the diagonal values are equal to '1' because a statement is always considered to be sorted into the same pile as itself.

Second, the individual sort matrices are added together to obtain a combined group similarity matrix. This matrix also has as many rows and columns as there are statements. Here, however, the value in the matrix for any pair of statements indicates *how many* people placed that pair of statements together in a pile regardless of what the pile meant to each person or what other statements were or were not in that pile. Values along the diagonal are equal to the number of people who sorted. Thus, in this square group similarity matrix, values can range from zero to the number of people who sorted. This final similarity matrix is con-



Binary Square Symmetric Similarity Matrix

	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 2. Procedure for computing the binary, symmetric similarity matrix for one person from the card sort.

sidered the relational structure of the conceptual domain because it provides information about how the participants grouped the statements. A high value in this matrix indicates that many of the participants put that pair of statements together in a pile and implies that the statements are conceptually similar in some way. A low value indicates that the statement pair was seldom put together in the same pile and implies that they are conceptually more distinct. There are many other ways than sorting to structure the conceptual domain, some of which are briefly described in Trochim and Linton (1986). The major advantages of the sorting procedure are that it is easily understandable by participants and that it takes little time to accomplish.

The second task in structuring the conceptual domain involves rating each statement on some dimension as described in the rating focus statement. Usually this rating is accomplished using a Likert-type response scale (e.g., 1-to-5 or 1-to-7 rating) to indicate how much importance, priority, effort or expected outcome is associated with each statement. For each statement one then obtains at least the arithmetic mean of the ratings and sometimes other descriptive statistical information.

Step 4: Representation of Statements

There are three steps involved in the way in which we typically represent the conceptual domain. First, we conduct an analysis which locates each statement as a separate point on a map (i.e., the point map). Statements which are closer to each other on this map were likely to have been sorted together more frequently; more distant statements on the map were in general sorted together less frequently. Second, we group or partition the statements on this map into clusters (i.e., the cluster map) which represent higher order conceptual groupings of the original set of statements. Finally, we can construct maps which overlay the averaged ratings either by point (i.e., the point rating map) or by cluster (i.e., the cluster rating map).

To accomplish the first step, the mapping process, we typically conduct a two-dimensional nonmetric multidimensional scaling of the similarity matrix obtained from Step 3. Nonmetric multidimensional scaling is a technique which takes a proximity matrix and represents it in any number of dimensions as distances between the original items in the matrix. A good introductory discussion of multidimensional scaling can be found in Kruskal and Wish (1978) and a more technical description of the algorithm which is used is given in Davison (1983).

A simple example of the principle which underlies multidimensional scaling can be given. If you were given a geographical map of the United States and asked to construct a table of distances between three major cities, say New York, Chicago, and Los Angeles, you could accomplish this fairly easily. You might take a ruler and measure the distances between each pair of cities and enter them into a 3 × 3 table of ruler-scale relative distances. However, if you were given only a table of distances between the three cities and were asked to draw a map which located the three cities on it as points in a way that fairly represented the relative distances in the table, the task would be slightly more difficult. You might begin by arbitrarily placing two points on a page to represent two of the cities and then try to draw in a third point so that its distances to the first two cities was proportionate to the distances given in the table. You would be able to accomplish this if the table consisted of three cities, but for more this task would become extremely complex. Multidimensional scaling is a multivariate analysis which accomplishes this task. It takes a table of similarities or distances and iteratively places points on a map so that the original table is as fairly represented as possible. In concept mapping, the multidimensional scaling analysis creates a map of points which represent the set of statements which were brainstormed based on the similarity matrix which resulted from the sorting task.

Typically, when multidimensional scaling analysis is conducted the analyst has to specify how many dimensions the set of points is to be fit into. If a one-dimensional solution is requested, all of the points will be arrayed along a single line. A two-dimensional solution places the set of points into a bivariate distribution which is suitable for plotting on an X-Y graph. The analyst could ask for any number of solutions from 1 to N − 1 dimensions. However, it is difficult to graph and interpret solutions which are higher than three-

dimensional easily. The literature on multidimensional scaling discusses this dimensionality issue extensively. One view is that the analyst should fit a number of solutions (e.g., one- to five-dimensional solutions) and examine diagnostic statistics to see whether a particular dimensional solution is compelling. This is analogous to examining J-plots of eigenvalues in factor analysis in order to decide on the number of factors. Another view suggests that in certain contexts automatic use of two-dimensional configurations might make sense. For instance, Kruskal and Wish (1978, p. 58) state that

> 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 3 or more dimensions.

In studies where we have examined other than two-dimensional solutions, we have almost universally found the two-dimensional solution to be acceptable, especially when coupled with cluster analysis as Kruskal and Wish (1978) suggest. Therefore, in concept mapping we usually use a two-dimensional multidimensional scaling analysis to map the brainstormed statements into a two-dimensional plot.

The second analysis which is conducted to represent the conceptual domain is called hierarchical cluster analysis (Anderberg, 1973; Everitt, 1980). This analysis is used to group individual statements on the map into clusters of statements which presumably reflect similar concepts. There are a wide variety of ways to conduct cluster analysis and there is considerable debate in the literature about the relative advantages of different methods. The discussion centers around ambiguity in the definition of the term "cluster." Everitt (1980) and Anderberg (1973) present more extensive discussions of this issue. We have tried a number of different cluster analysis approaches. Originally, we used an oblique principle components factor analysis approach to hierarchical cluster analysis where the input for the analysis consisted of the similarity matrix. The problem with this approach was that it often led to results which did not visually correspond with the way in which multidimensional scaling mapped the points. This is because differences in the two algorithms (i.e., multidimensional scaling and cluster analysis), when applied to the same similarity matrix, sometimes meant that points which were close to each other on the map were placed in separate clusters by the cluster analysis. These results were hard to interpret and seemed to give equal weight to multidimensional scaling and cluster analysis. Instead, it makes sense to view the mathematical basis for multidimensional scaling as stronger than basis for cluster analysis and, accordingly, to rely on the multidimensional scaling rather than cluster analysis to depict the basic inter-statement conceptual similarities. What we wanted was a cluster analysis which grouped or partitioned the statements on the map as they were placed by multidimensional scaling. We found that this could be accomplished by using the X-Y multidimensional scaling coordinate values for each point (rather than the original similarity matrix) as input to the cluster analysis. In addition, we also found that Ward's algorithm for cluster analysis (Everitt, 1980) generally gave more sensible and interpretable solutions than other approaches (e.g., single linkage, centroid). Therefore we have moved to an approach which uses Ward's hierarchical cluster analysis on the X-Y coordinate data obtained from multidimensional scaling as the standard procedure. This in effect partitions the multidimensional scaling map into any number of clusters.

Just as deciding on the number of dimensions is an essential issue for multidimensional scaling analysis, deciding on the number of clusters is essential for cluster analysis. All hierarchical cluster analysis procedures give as many possible cluster solutions as there are statements. In principle, these clustering methods begin by considering each statement to be its own cluster (i.e., an N-cluster solution). At each stage in the analysis, the algorithm combines two clusters until, at the end, all of the statements are in a single cluster. The task for the analyst is to decide how many clusters the statements should be grouped into for the final solution. There is no simple way to accomplish this task. Essentially, the analyst must use discretion in examining different cluster solutions to decide on which makes sense for the case at hand. Usually, assuming a set of a hundred or fewer statements, we begin by looking at all cluster solutions from about 20 to 3 clusters. Each time the analysis moves from one cluster level to the next lowest (e.g., from 13 to 12 clusters) we examine which statements were grouped together at that step and attempt to decide whether that grouping makes sense for the statements in the conceptualization. In examining different cluster solutions we have found it useful to use a cluster tree which shows pictorially all possible cluster solutions and mergers. In general, we attempt to decide on a cluster solution which, if anything, errs on the side of more clusters than fewer. Clearly this is a task which requires discretion on the part of the analyst and it would be ideal if we could involve the participants directly in this decision making process, but as of yet we have not determined an easy way to accomplish this kind of task within a group process.

Our experience shows that, in general, the cluster analysis results are less interpretable than the results from multidimensional scaling. The cluster analysis is viewed as suggestive and, in some cases, one may want

to "visually adjust" the clusters into more sensibly interpretable partitions of the multidimensional space. The key operative rule here would be to maintain the integrity of the multidimensional scaling results, that is, try to achieve a clustering solution which does not allow any overlapping clusters (e.g., a true partitioning of the space).

Once we have conducted the multidimensional scaling and cluster analysis, we are able to generate a point and a cluster map. The final analysis involves obtaining average ratings across participants for each statement and for each cluster. These can then be overlayed graphically on the maps to produce the point rating map and the cluster rating map as will be shown later.

We have several products at the end of the representation step. First, we have the two-dimensional point or statement map which locates each of the brainstormed statements as a point. Next to each point we place the number of the statement so that participants can identify each point as a statement. Second, we have a cluster map which shows how the cluster analysis grouped the points. Third, we have the point rating map which shows the average ratings for each statement on the point map. Finally, we also have the cluster rating map which shows the average rating for each cluster on the cluster map. This information forms the basis of the interpretation in the next step.

Step 5: Interpretation of Maps

To interpret the conceptualization, we usually assemble a specific set of materials and follow a specific sequence of steps—a process which has been worked out largely on the basis of our experiences on many different projects. The materials consist of:

1. *The Statement List.* The original list of brainstormed statements, each of which is shown with an identifying number.
2. *The Cluster List.* A listing of the statements as they were grouped into clusters by the cluster analysis.
3. *The Point Map.* The numbered point map which shows the statements as they were placed by multidimensional scaling.
4. *The Cluster Map.* The cluster map which shows how statements were grouped by the cluster analysis.
5. *The Point Rating Map.* The numbered point map with average statement ratings overlayed.
6. *The Cluster Rating Map.* The cluster map with average cluster ratings overlayed.

Notice that there are four different types of maps here. Which of them is *the* concept map? In fact, they are *all* concept maps. Each of these maps tells us something about the major ideas and how they are interrelated. Each of them emphasizes a different part of the conceptual information. While the maps are distinctly different ways of portraying or representing the conceptual structure (and consequently, different names are used to distinguish them), it is important to remember that they are all related to each other and are simply reflecting different sides of the same underlying conceptual phenomenon. In the remainder of this paper, if the type of concept map is not specified, it is fair to assume that the discussion pertains to the cluster map because that is usually the most directly interpretable map.

The facilitator begins by giving the group the original set of brainstormed statements (for the York County example, the statement list is shown in Table 1 above) and recalling that the statements were generated by the group in the brainstorming session. Participants are then reminded that they grouped these statements into piles and told that the individual sortings were combined for the entire group. The statements as they were grouped by the cluster analysis (the cluster list) are then presented. For the York County study, the cluster listing is given in Table 2.

Each participant is asked to read through the set of statements for each cluster and come up with a short phrase or word which seems to describe or name the set of statements as a cluster. This is analogous to naming factors in factor analysis. When each person has a tentative name for each cluster, the group works cluster-by-cluster in an attempt to achieve group agreement on an acceptable name for each cluster. This is an often interesting negotiating task. When each person in turn gives their name for a certain cluster, the group can often readily see a consensus which exists. For some clusters, the group may have difficulty in arriving at a single name. This is because the statements in that cluster might actually contain several different ideas and, had a higher cluster solution been selected, the statements would have been subdivided into subclusters. In these cases, the facilitator might suggest that the group use a hybrid name, perhaps by combining titles from several individuals. In any event, the group is told that these names are tentative and may be revised later.

When the group has reached consensus on the names for each cluster, they are presented with the numbered point map. The York County map is shown in Figure 3.

They are told that the analysis placed all of the statements on the map in such a way that statements which were piled together frequently should be closer to each other on the map than statements which were not piled together frequently. Usually it is a good idea to give them a few minutes to identify a few statements on the map which are close together and examine the wording of those statements on the original brainstormed statement list as a way to reinforce the notion that the analysis is placing the statements sensibly. When they have become familiar with this numbered point map, they

TABLE 2
STATEMENTS GROUPED BY CLUSTER, YORK COUNTY ELDERLY PROJECT

Cluster 1
- 1. lack of public transportation
- 13. inadequate home care services
- 2. lack of money
- 75. in-home visits by physicians
- 78. clergy who can make home visits

Cluster 2
- 3. high cost of medication
- 4. high cost of health care
- 51. long-term care insurance
- 52. Medicare coverage being gradually "squeezed down"
- 91. high cost of home nursing care
- 53. lack of state and federal funding commitment to elderly

Cluster 3
- 24. lack of funding for community care
- 32. Medicare limitations
- 36. housing costs
- 70. access issues for disabled elderly
- 76. clergy that are aware of community resources

Cluster 4
- 26. confusing social services system
- 29. duplication of services
- 89. better understanding of elderly by legal profession
- 50. awareness of community resources

Cluster 5
- 14. lack of nursing homes
- 49. need for respite for caregivers
- 94. lack of support for family caregivers
- 18. emergency supervised housing
- 37. different housing alternatives
- 83. infantilization of elderly by providers

Cluster 6
- 27. exploitation
- 28. physical abuse and neglect by caretakers

Cluster 7
- 30. lack of outreach
- 33. nutrition services
- 34. expenses of maintaining family contacts

Cluster 8
- 25. lack of expertise to treat mentally ill elderly
- 95. law enforcement intervention when elderly refuse to be hospitalized
- 73. placement outside of home
- 39. over-medication by physicians
- 60. assistance with managing finances

Cluster 9
- 48. deterioration of family life
- 90. distress about the family issues of their children
- 67. care for elderly persons with dementia
- 93. depletion of resources by children

Cluster 10
- 5. stigmatization of elderly persons
- 61. job discrimination
- 66. premature categorization as "old" or elderly

Cluster 11
- 44. continuation of vocational satisfaction
- 69. improved use of educational opportunities
- 62. lack of education on life changes
- 74. means to maintain independence
- 65. how do some elderly function so well? (exceptional elderly)
- 92. role models for the elderly
- 64. desire to pass on traditions and knowledge/experiences
- 72. more diverse social opportunities

Cluster 12
- 35. need for political advocacy
- 58. desire for access to educational/recreational programs
- 63. exclusion of elderly as an educational resource
- 71. accessible education through creative programming
- 77. improved public awareness and support of elderly needs
- 43. expanding (exploding) elderly population

Cluster 13
- 11. late life role changes
- 45. lack of preparation for aging (including retirement)
- 57. feeling of being excluded by other generations or groups
- 21. inability to cope with change
- 85. giving elderly persons the right to choose to die
- 46. issues around self-worth and self-esteem

Cluster 14
- 16. loss of recreation
- 19. lack of socialization
- 59. desire to be treated like everybody else
- 79. understanding and respect for sexuality
- 41. need for independence
- 47. changing role in the community

Cluster 15
- 6. isolation
- 23. death of spouse
- 68. confusion between aging and illness
- 17. concern for mental wellness
- 54. fear of prolonged dying
- 86. fear of being abandoned
- 20. fears about not being able to protect themselves

Cluster 16
- 9. deterioration of mental capacity
- 38. over self-medication
- 12. depression
- 80. fear of being crazy
- 81. suicide among the elderly
- 55. concern over being a burden
- 56. feeling of uselessness

Cluster 17
- 7. dependency—interpersonal
- 8. loss of close support
- 87. loss of legal control
- 40. inability to understand or read medication instructions
- 42. lack of awareness in mixing drugs and alcohol
- 82. alcoholism among the elderly

Cluster 18
- 10. physical disabilities
- 88. loss of financial control
- 15. decreased functional status
- 31. losing your driver's license
- 22. effects of the weather on the elderly
- 84. an institutional end (decline and death)

[illegible] that the analysis also organized the points into [illegible] on the list of clustered statements [illegible]amed. The cluster map is presented [illegible] are shown that the map portrays visually the exact same clustering which they just looked at on the cluster list. They are then asked to write the cluster names which the group arrived at next to the appropriate cluster on the cluster map. They are then asked

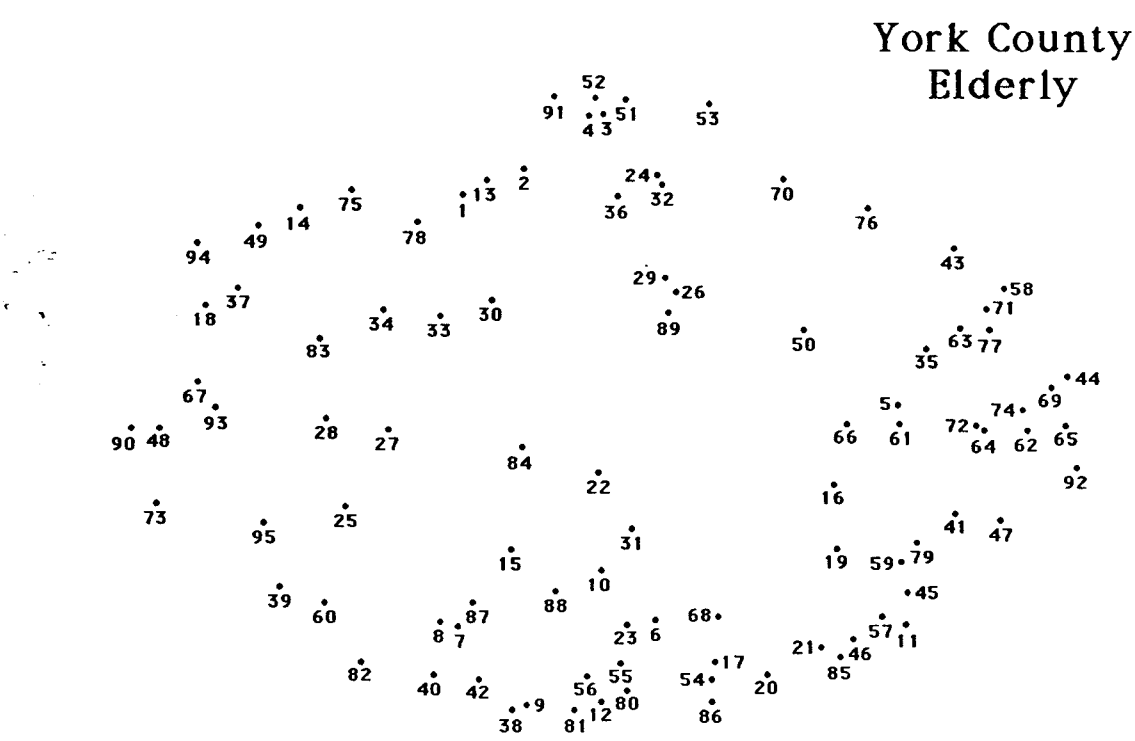


Figure 3. Numbered point map for the York County Elderly project.

to examine this named cluster map to see whether it makes any sense. The facilitator should remind participants that in general, clusters which are closer together on the cluster map should be more similar conceptually than clusters which are farther apart and ask them to assess whether this seems to be true or not. Participants might even begin at some point on the map and, thinking of a geographic map, "take a trip" across the map reading each cluster in turn to see whether or not the visual structure makes any sense. For the York County example, the named cluster map is given in Figure 4.

The participants are then asked to see whether there are any sensible groups or clusters of clusters. Usually, the group is able to perceive several major regions. These are discussed and partitions are drawn on the map to indicate the different regions. Just as in naming the clusters, the group then attempts to arrive at a consensus concerning names for these regions. In the York County study, people were satisfied with the clustering arrangement and did not wish to define any regions.

This final named cluster map constitutes the conceptual framework and the basic result of the concept mapping process. The facilitator should remind the participants that this final map is their own product. It was entirely based on statements which they generated in their own words and which they grouped. The labels on the map represent categories which they named. While in general the computer analysis will yield sensible final maps, the group should feel free to change or rearrange the final map until it makes sense for them and for the conceptualization task at hand. At this point it is useful for the facilitator to engage the participants in a general discussion about what the map tells them about their ideas for evaluation or planning.

If ratings were done in the structuring step, the facilitator then presents the point rating and cluster rating maps. The point rating map for the York County Study is given in Figure 5, while the cluster rating map is shown in Figure 6. Participants examine these and attempt to determine whether they make sense and what they imply about the ideas which underlie their evaluation or planning task.

In the York County study some interesting insights arose from the interpretation process. If we look at either of the cluster maps (Figure 4 or 6) we can see that the clusters on the right side of the map—Personal Growth and Education, Stereotyping, Socialization Needs, and Political Strength and Advocacy—tend to be the types of issues of most concern to the "well elderly," people who are not ill or institutionalized. As we move counter-clockwise on the map, the clusters on the top represent concerns of those who are becoming ill or are in need of service, those on the left are most relevant for people who are already ill and/or in need of intensive home or center-based care, and those on the bottom pertain most to persons who are severely ill or dying. The group perceived this counter-clockwise cycle as a good description of their implicit theory of the aging process. Furthermore, they believed that most

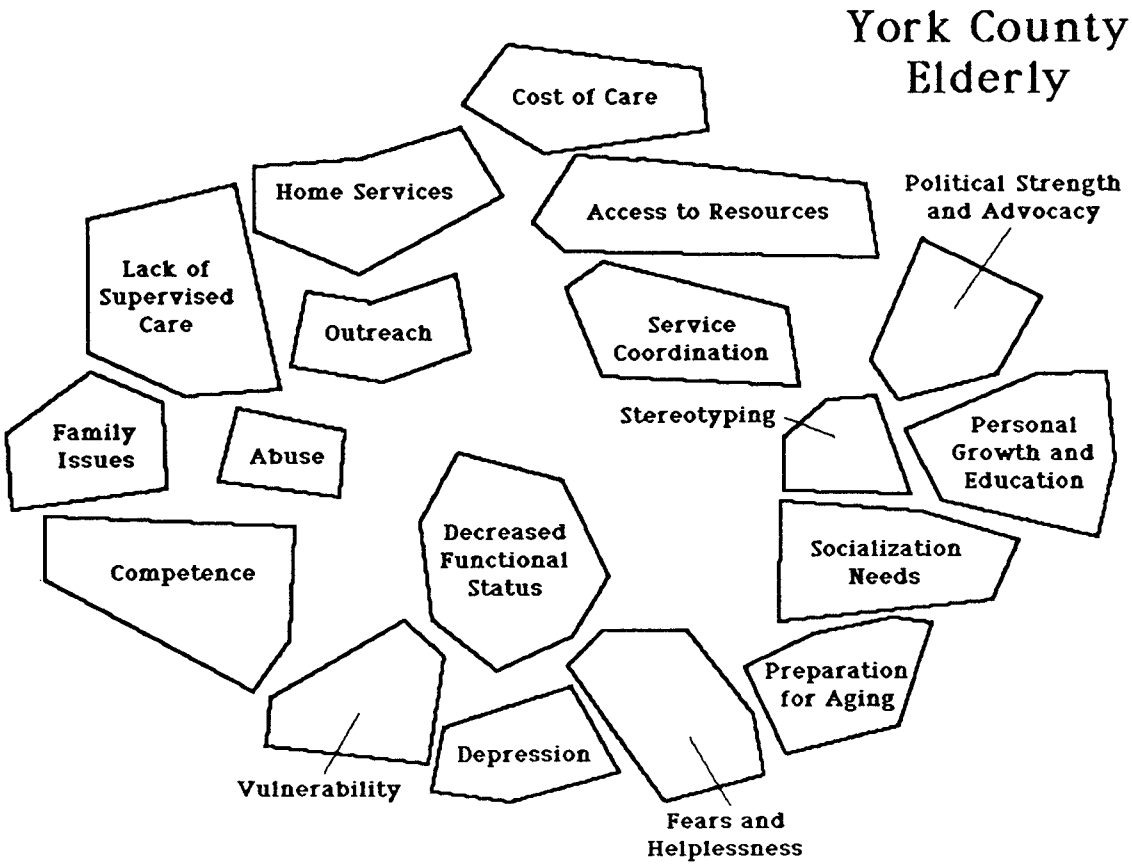


Figure 4. Named cluster map for the York County Elderly project.

of the political strength and advocacy work that exists around aging issues tends to be done by the "well elderly" as the map implies. This led them to discuss the desirability of working within the community to encourage the well elderly to perceive their position within the entire aging cycle and begin to engage in more active advocacy for the full range of concerns which the map describes. Thus, the map provided a foundation for an approach to the elderly which addresses major concerns and emphasizes the more active involvement especially of the well elderly.

Step 6: Utilization of Maps

At this point in the process we turn our attention back to the original reason for conducting the structured conceptualization. The group discusses how the final concept map might be used to enhance either the planning or evaluation effort. The uses of the map are limited only by the creativity and motivation of the group. A number of straightforward applications suggest themselves. For instance, if the conceptualization was done as the basis for planning, the final map might be used for structuring the subsequent planning effort. The planning group might use it for dividing up into subgroups or task forces, each of which is assigned a specific cluster or region. Each task group could then examine issues like: the organizational budget allocation for each cluster, how organizational personnel are distributed within each cluster, how important each cluster is relative to the others, what resources might be brought to bear in addressing each cluster, what level of competition exists from other organizations providing services in each cluster, and so on. The task forces can use the individual statements within a cluster as cues or prompts concerning what they should consider specifically within each cluster. One major advantage of having the concept map is that the results of these task force investigations can often be usefully displayed directly on the concept map as were the priority ratings for the York County example in Figure 5 or 6. Thus, any number of maps can be created showing such variables as budget allocations, staff effort, or degree of need, and displayed by individual statement and/or by cluster.

For planning purposes, the concept map can also be used as the framework for an outline of a planning report. Regional headings would constitute the highest level of indentation for the outline, clusters would be subheaded within their appropriate regions, individual statements could be subheadings within clusters, and

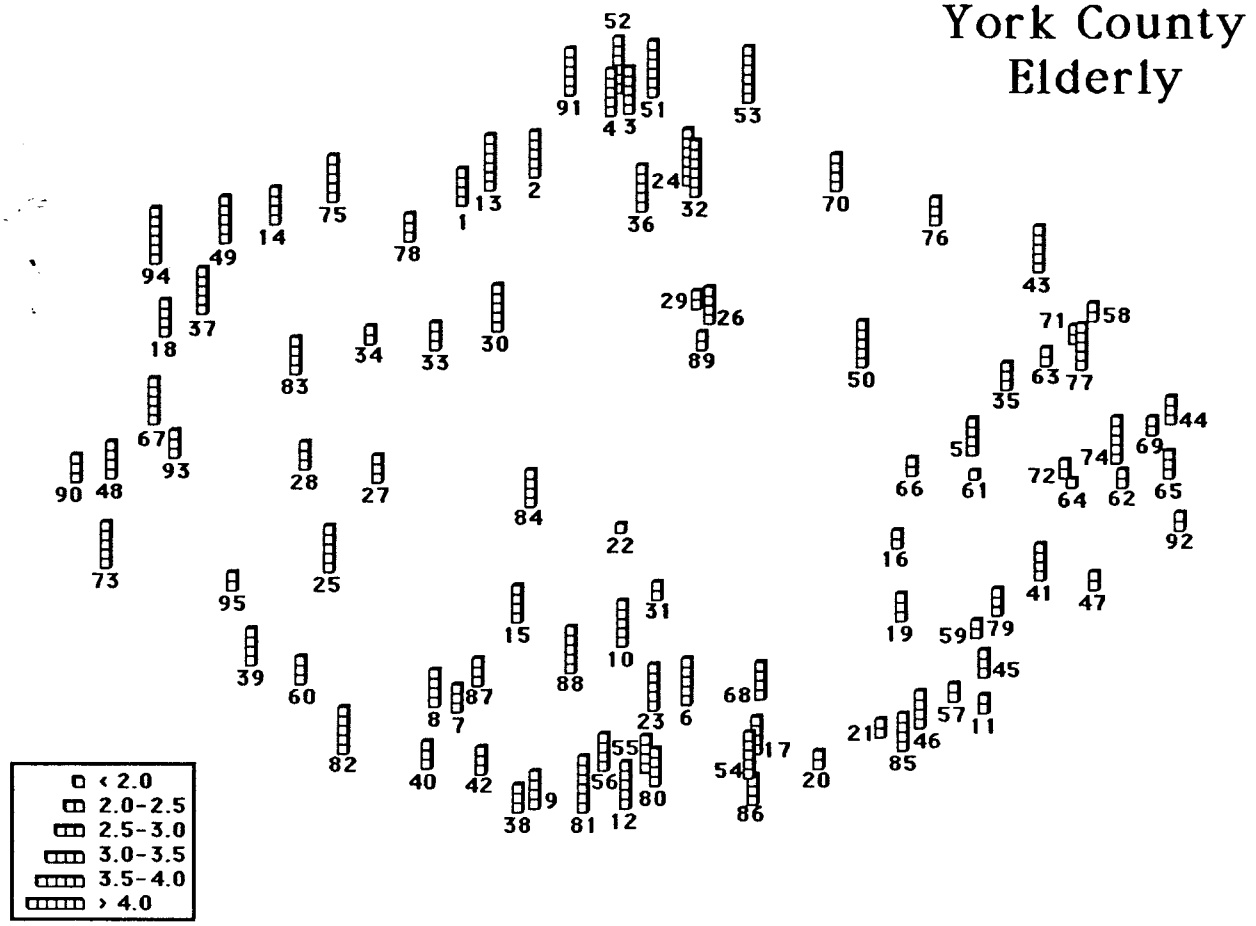


Figure 5. Point rating map for the York County Elderly project.

any statement-level information relevant to planning could be subheaded within this structure. Thus for planning, the concept map provides a framework for understanding important issues in a way which enables sensible pictorial and outline representations.

The concept map is also extremely useful in evaluation contexts. Here, its utilization depends on what the focus was for the conceptualization. If the focus was on planning a program or service which would then be evaluated, the concept map can act as an organizing device for operationalizing and implementing the program. For instance, if the program is a training program in a human service agency, the training can be constructed based on the concept map with different training sessions designed to address each cluster and the individual brainstormed statements acting as cues for what kinds of information should be covered in each session. The concept map is the framework for the program construct and can form the basis of a process evaluation of the program. In this case or in the case where the focus of the conceptualization was on the outcomes of some program, the concept map can guide measurement development. Each cluster can be viewed as a measurement construct and the individual statements can suggest specific operationalizations of measures within constructs. For instance, if the group wished to develop a questionnaire, they could use the concept map by having each cluster represented with questions on the questionnaire. Furthermore, the original brainstormed statements might provide question prompts which either directly or with some revision could be included on the questionnaire along with some rating response format. Alternatively, if a more multimethod approach to measurement was desired, the group could make sure that within each cluster several different types of measures were constructed to reflect the cluster. The exciting prospect here is that the concept map provides a useful way to operationalize the multitrait-multimethod approach to measurement which was outlined by Campbell and Fiske (1959) and is described in greater detail in the paper by Davis (this volume). In this example, the concept map represents the group's theoretical expectations about how the major measurement constructs are conceptually interrelated. From the

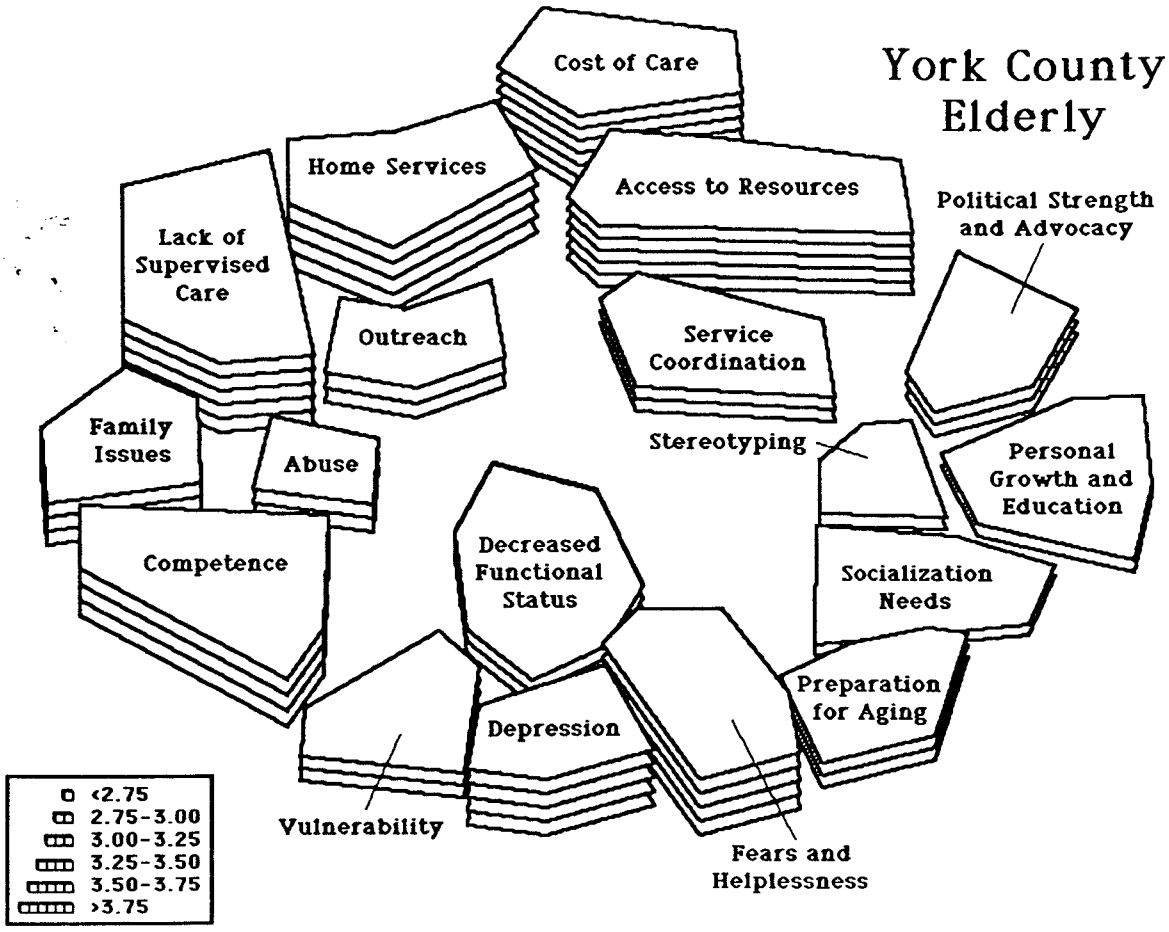


Figure 6. Cluster rating map for the York County Elderly project.

concept map we can predict the rank order which we expect in the correlations between measures. These expectations (the theoretical pattern) could then be directly compared with a matrix of correlations as obtained in the study (the observed pattern) and the degree to which the two match can constitute evidence for the construct validity of the measures. This "pattern matching" approach to construct validity is discussed in detail in Trochim (1985; in press).

A number of papers in this volume illustrate the use of structured conceptualization for evaluation. For instance, Valentine (this volume) used concept mapping to construct an instrument which could be used to assess caring in a nursing context. Galvin (this volume) used concept mapping to describe the central concepts in a Big Brother/Big Sister program, constructed an instrument from the map, and used the instrument to explore how well the program appeared to be performing.

COMPUTER PROGRAMS

An introduction to concept mapping would be incomplete without some consideration of the computer programs which can be used to accomplish this process. Essentially, there are two options: use some combination of standard general-purpose word processing and statistics packages; or, use the computer package which was designed by the author specifically for accomplishing concept mapping. Each is discussed in turn.

Using General-Purpose Software Packages

When using available general-purpose software packages, the analyst will have to be prepared to experiment with different processing options until a suitable procedure can be constructed. Minimally, it is desirable to have a good word processing program; a statistics package which has routines for multidimensional scaling and cluster analysis, and which has fairly flexible data manipulation capabilities; and, a graphics program to plot the final maps.

On a mainframe computer, brainstormed statements can be entered into any standard editor. They would then need to be formatted and printed onto cards, la-

bels, or sheets in a manner which allows the analyst to assemble them into sorting decks. In addition, the statements would need some minor formatting in order to produce an instrument for the rating task. Multidimensional scaling and cluster analysis are available in most large scale statistical systems including SAS and SPSSx. The major difficulty with this approach will involve the entry and manipulation of the sort data. The sort results could be entered into a data matrix using a standard editor where each participant constitutes a single record (row) and there are as many variables (columns) as there are statements. The analyst would arbitrarily assign a unique number to each pile of sorted statements and enter this number in the appropriate row and column. There are two main difficulties with this procedure, both of which are resolvable. First, since the data for any participant will be entered in the order of the statements, the sort data will have to be coded. That is, the analyst will have to set up a coding sheet with the statements numbered sequentially and enter the pile number next to each statement before entering the data. Second, and perhaps more troublesome, the analyst will have to write a computer program which takes the sort data matrix as input, constructs the binary square similarity matrix for each person and then adds these to get the group similarity matrix. This matrix can then be ported to the statistics package for input to the analysis. On the graphics side, it is possible to use the plotting routines available with statistical packages to obtain the point map. However, it may be difficult to obtain automatically the cluster drawings and any 3-D or pseudo-3-D plots which overlay ratings as shown in Figures 5 and 6. Listings of statements by cluster will almost certainly have to be accomplished either by re-entering the statements in cluster order or by editing the original file.

The concept mapping analyses can also be accomplished on microcomputers, although there are some trade-offs involved. For instance, there are few microcomputer implementations of multidimensional scaling available, although there are many graphics packages which might be easier to use than mainframe counterparts to produce the maps. Again, one could use virtually any word processing program for entering the statements (e.g., Microsoft WORD, WordPerfect). As with the mainframe option, the statements would need to be formatted for printing as sorting decks and as an instrument for the rating task. The SYSTAT program, available for both MS/DOS and Macintosh environments is one of the few microcomputer programs at this time which includes both multidimensional scaling and cluster analysis (although the cluster analysis package does not include Ward's algorithm). As with the mainframe option, the analyst will have to write a program to construct the group similarity matrix. The maps can be produced using a combination of plotting and painting programs. For instance, one might use CricketGraph on the MAC (or even the SYSTAT Graph option) to generate the x-y point plot and then load that picture into a painting program (e.g., SuperPaint, FullPaint, CANVAS) to manually draw cluster boundaries and construct pseudo-3-D rating plots like those in Figures 5 and 6. Again, lists of statements by cluster will have to be constructed by editing the original brainstormed statement file.

Using The CONCEPT SYSTEM

Because of the inconvenience of using general-purpose programs for accomplishing concept mapping, the author has written a computer program, called *The Concept System* specifically to accomplish this task. The program is available for both the MS/DOS and Macintosh microcomputer environments. The program is interactive and has separate menu-selected options for: entering the brainstormed statements; printing decks of cards or pre-formatted rating sheets; entering sort data (the data can be entered by cluster simply by typing in the ID numbers of the statements in any given cluster); entering the rating data; conducting the analysis (including construction of the similarity matrix, multidimensional scaling, cluster analysis, and averaging of ratings); and, interactive graphing of the results in a wide variety of ways to produce any of the maps discussed above. The Concept System allows the user to examine interactively any possible clustering solution, print a cluster tree and lists of statements by any selected number of clusters. The program limits the user to no more than 100 statements in any concept mapping project. Anyone interested in obtaining information about The Concept System program may do so by contacting the author directly. Virtually all of the projects reported in this volume utilized The Concept System to compute maps. While this is undoubtedly the easier way to accomplish concept mapping because the program was written specifically for this purpose, it is certainly also feasible to do so using available general-purpose programs as well.

CONCLUSIONS

Concept mapping of the type described here is designed to bring order to a task which is often extremely difficult for groups or organizations to accomplish. The process has several distinct advantages (many of which are discussed in greater detail in the other papers in this volume). First, it encourages the participant group to stay on task and to lay out relatively quickly a framework for a planning or evaluation study. Second, it expresses the conceptual framework in the language of the participants rather than in terms of the evaluator's

or planner's language or the language of social science theorizing. Third, it results in a graphic representation which at a glance shows all of the major ideas and their interrelationships. Fourth, this graphic product is comprehensible to all of the participants and can be presented to other audiences relatively easily. Finally, we have observed over many concept mapping projects that one of the major effects of the process is that it appears to increase group cohesiveness and morale. Especially in groups which have previously tried to accomplish conceptualizing through committee discussions, we have found that they readily appreciate the structure of the process and the ease with which it produces an interpretable starting point for subsequent evaluation or planning work.

This concept mapping process is by no means the only way in which group conceptualization can be accomplished nor it is necessarily the best way for any given situation. In situations where a group can achieve consensus relatively easily on its own or where a pictorial representation of its thinking is not desired or deemed useful, this approach would not be recommended.

As the other papers in this volume will show, we are only just beginning our efforts to devise sensible group conceptualization processes. There are still many methodological issues which need to be explored in order to improve the process described here, and that work is continuing. However, as it now stands, the concept mapping process is a useful procedure which helps a group to focus on the conceptualization task, and results in an easily understandable pictorial representation of their thinking which emerges from a group process which insures input from all participants.

REFERENCES

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

BICKMAN, L. (Ed.). (1986). Using program theory in evaluation. *New directions for program evaluation*. San Francisco: Jossey-Bass.

CAMPBELL, D.T., & FISKE, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

CHEN, H.T., & ROSSI, P.H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283-302.

CHEN, H.T., & ROSSI, P.H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10*, 95-103.

CORDRAY, D.S. (1986). Quasi-experimental analysis: A mixture of methods and judgment. In W. Trochim (Ed.), Advances in quasi-experimental design and analysis. *New directions in program evaluation*. San Francisco: Jossey-Bass.

DAVISON, M.L. (1983). *Multidimensional scaling*. New York: John Wiley and Sons.

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

EINHORN, H.J., & HOGARTH, R.M. (1986). Judging probable cause. *Psychological Bulletin, 99*, 1, 3-19.

EVERITT, B. (1980). *Cluster analysis* (2nd ed.). New York: Halsted Press, A Division of John Wiley and Sons.

KRIPPENDORF, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.

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

MARK, M.M. (1986). Validity typologies and the logic and practice of quasi-experimentation. In W. Trochim (Ed.), Advances in quasi-experimental design and analysis. *New directions in program evaluation*. San Francisco: Jossey-Bass.

NOVAK. J.D., & GOWIN, D.B. (1984). *Learning how to learn*. Cambridge, England: Cambridge University Press.

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

RICO, G.L. (1983). *Writing the natural way: Using right-brain techniques to release your expressive powers*. Los Angeles, CA: J.P. Tarcher.

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

ROSSI, P.H., & CHEN, H. (in press). Issues and overview of the theory-driven approach. *Evaluation and Program Planning*.

SHADISH, W.R., COOK, T.D., & HOUTS, A.C. (1986). Quasi-experimentation in a critical multiplist mode. In W. Trochim (Ed.), Advances in quasi-experimental design and analysis. *New directions in program evaluation*. San Francisco: Jossey-Bass.

STONE, P.J., DUNPHY, D.C., SMITH, M.S., & OGILVIE, D.M. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge: The Massachusetts Institute of Technology.

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

TROCHIM, W. (in press). Pattern matching and program theory. *Evaluation and Program Planning*.

TROCHIM, W., & LINTON, R. (1986).Conceptualization for evaluation and planning. *Evaluation and Program Planning, 9*, 289-308.

WRIGHTSON, M. (1976). The documentary coding method. In R. Axelrod (Ed.), *The structure of decision: The cognitive maps of political elites*. Princeton, NJ: The Princeton University Press.

Evaluation and Program Planning, Vol. 12, pp. 17–23, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

CONTRIBUTIONS TO THE THEORY OF CARE

KATHLEEN VALENTINE

Cornell University

ABSTRACT

This paper discusses how structured conceptualization, conducted as one phase of a larger study, was used to conceptualize "caring" in a nursing context. Variations of the structured conceptualization process were conducted with three different groups; nurses, patients, and nurse theorist/researchers and a separate concept map was derived for each group. The maps are discussed along with implications for the development of a theory of caring and consideration of the advantages and disadvantages of the methodology.

Currently, theorists of caring call for naturalistic inquiry for theory building; while health care managers require rational information for decision making. The results of naturalistic inquiry are often expressed in words, while the results of rationalistic inquiry are often expressed in numbers. The focus of this paper is to describe one investigation of caring between nurses and patients which seeks to meet the information needs of managers while recognizing the current state of theory and research about caring. Trochim's structured conceptualization technique (Trochim, this volume; Trochim & Linton, 1986) is useful in obtaining multiple perspectives for the explication of program theory. In addition, the results can be used to develop quantitative measures of caring. Thus, it helps to serve the information needs of both theorists and managers.

The main focus of this paper is to discuss how the structured conceptualization activities, conducted as one phase of a larger study, helped to form a conceptualization of caring. A secondary consideration involves examining the implications of the concept maps for the eventual development of a questionnaire to measure caring.

CONTEXT OF CARING

Human services in the 1980s have had to struggle with the tension between economic and humanistic societal issues. As the world economy moves from production to service, the call for rational management of scarce resources increases. Scientific management practices require rational information for decision making and cost-benefit, and cost-effectiveness approaches are often used as aids for such decision making. In the health care industry, cost containment efforts have predominated policy and reimbursement decisions. Simultaneously, the industry is expected to increase or at least maintain access and quality of health care, and distribute the service equitably. Thus, the industry must address the tension between economic pressures and humanistic values.

One expression of humanistic values is caring. Caring or care is a term which is used frequently and taken for granted as an innate human quality. In health care, caring is considered a central component of health services. For nursing, it is considered "the essence and central focus of nursing practice" (Leininger, 1984). Though caring is the essence and central focus of nursing, research about caring is in its very early stages, and theories of professional nursing care are emerging (Leininger 1981, 1984, and Watson, 1979, 1985). To date, there is no universally agreed upon definition of caring, or professional nurse caring. In order to quantify a variable, one must first be able to define it. This presents a challenge in today's competitive health care environment which values rational information for decision making. In order for care to be given priority, it must be expressed in terms comparable to other fac-

Requests for reprints should be sent to Kathleen Valentine, 15 Marvin Street, Clinton, NY 13323.

tors used in resource allocation decisions. The problem is, that the theory and research about caring does not currently meet the information needs of managers who frequently desire quantification and objectification of relevant variables for decision making.

What is needed is an investigation of caring which recognizes the current state of caring theory and research yet communicates effectively to managers about the necessary resources to insure its presence. Trochim's (this volume; Trochim and Linton, 1986) structured conceptualization process is a useful tool for collecting and analyzing different perspectives about caring. The concept maps generated through multidimensional scaling and cluster analysis serve as the basis for discussion and emergence of a grounded theory of caring which is particular to a specific setting.

METHOD

Reasons for Choosing Structured Conceptualization Method

This method was chosen for four reasons:

1. To provide a concrete representation for the abstract concept of caring which is useful for the discussion, understanding, and communication of that concept.
2. To develop a grounded definition of caring which was based on both qualitative and quantitative methods.
3. To provide the background information for development of a Likert-type questionnaire to measure caring.
4. To provide an opportunity for greater exploration of construct validity of caring through pattern matching.

The structured conceptualization process allows one to make abstract concepts concrete through visual representations. This picture then helps people to focus discussion on the elements which make up the concept, and the interrelationships between the elements of the concept. This provides a dynamic understanding of the concept rather than a mere listing of its elements. This is especially useful when talking about human processes which by their very nature are dynamic. The concept maps provide a framework from which other relationships can be discussed. Participants often express that the maps do represent some underlying relationships which they had previously considered but which had not been fully expressed. The quantitative nature of the multidimensional scaling and cluster analysis lend strength to the interpretive discussions of the concept map because the participants have confidence that the maps represent measurement of some common ideas. This sense-making phase of the process also allows for discussion of other related aspects of the concept which are not represented by the maps. Thus, an integrated grounded definition of caring using mixed methods can increase confidence that the conceptual domain of interest is being adequately represented.

This grounded conceptualization of caring and the examination of interrelationships can serve as the basis for questionnaire item development. Items can be developed which represent both the unique characteristics of caring found on different maps, as well as the overlapping concepts of caring. This gives greater confidence in the instrument's validity.

Aside from the validity of the instrument, structured conceptualization can contribute to construct validity through pattern matching processes (Trochim, 1985). Do the theoretical relationships expressed in the concept maps remain viable when compared with obtained measures? If they do match, then theoretical conceptualizations are strengthened. If they do not match, then there may be problems with the theoretical conceptualization and/or the measurement process. This feedback process between theory and practice is aided by the concrete nature of structured conceptualization process in which one can "see" the patterns and relationships.

Concept Mapping Processes

Variations of the structured conceptualization process were conducted with three different groups; nurses, patients, and nurse theorist/researchers as shown in Table 1. Each of these processes is described in turn.

Nurses. This study focuses on caring which occurs between nurses and patients in the acute care hospital setting. Thus, a sample of acute care hospital nurses were asked the question, "When you think about the patient in the hospital setting, what do you think of as caring?" This question was posed to small informal groups of nurses of about 3–6 people each. Approxi-

TABLE 1
STEPS IN CONCEPT MAPPING PROCESS FOR THE THREE PARTICIPANT GROUPS

	Concept Mapping Steps		
Participant Group	Generation	Sorting/Rating	Interpretation
Nurses	Brainstorming	Yes	Yes
Patients	No, used above with revisions	Yes	Yes
Nurse Theorists/ Researchers	No, used items from literature	Yes	Yes

mately 40 people were involved in generating their ideas of caring. Without editing (except for absolute duplication of items), the entities which were generated were put on 3 by 5 cards. Members of the nursing council, who represent a cross section of nurses from the hospital, then sorted these cards in "any way which made sense to them." They also rated each of the 80 entities on a scale of 1-5 with 1 being least central to caring, and 5 being most central to caring. This same sorting and rating process was also done with a group of nurse managers. The data from the two groups of nurses were then combined and one concept map for nurses was computed. The map was then interpreted in each of two meetings, one with nursing staff, and one with nurse managers. The results were then combined into one interpreted map for nurses.

Patients. Eleven female members of a hospital sponsored health information group participated as volunteers for research about caring. This sample was chosen as representative of the patients who will be administered a questionnaire about caring during a later phase of the study. That group of patients will be women hospitalized for hysterectomies. These female volunteers were given the same eighty items which the nurses had generated. They were asked to review the items to add any that they felt were missing. No items were added; however, clarification of the meaning for some of the items was provided. The women then sorted and rated the items in the same manner as the nurses. The group then convened to interpret the map and discuss the results.

Nurse Theorists and Researchers of Caring. This group meets annually for a national conference on Care and Caring at which the latest research and theoretical concepts are shared. Twelve nurse researchers and theorists participated in the process of sorting and rating items on caring. Seventy-nine items were taken from the literature on caring by the investigator. The data were analyzed and a concept map generated. An initial interpretation of the map was obtained from some nurse theorists. Further interpretation will be obtained during the next annual meeting of this group.

Some Methodological Concerns

The structured conceptualization process takes time to complete in its entirety from generation of entities through map interpretation. Only the nurse group participated fully in all three phases (see Table 1). Patients and nurse theorists were less available for all three phases, due to access and time constraints. Therefore, it was decided to use data sets which those participants did not generate. However, they did sort, rate, and interpret the maps. The decision to eliminate the generation step may affect confidence that the items represent the conceptual domain for those participants. However, the effects of this are minimized by allowing for revision of the data set through addition of other items. The nurse researchers and theorists did add some items; the patient group did not add items.

RESULTS

Interpretation of Concept Maps

Nurses. Figure 1 is the concept map for nurses, as they interpreted it. In interpretation, it is assumed that clusters which are closer in distance are also more similar. The higher the average rating for items in that cluster, the more central are those items to the concept of caring and the more layers the cluster has on the map.

Cluster 3 includes positive innate affective human qualities which are especially important qualities for a nurse to have. Clusters 7, 6, and 5 include cognitive aspects of caring as well as the affective qualities found in cluster 3. These three clusters also include the element of action, that is expression of caring through action. Cluster 7 demonstrates this through "creating a climate in which a therapeutic relationship can be developed." Cluster 6 is "providing personalized care to a patient while respecting that person's autonomy." Cluster 5 includes those actions which "enable a person toward independence," helping a person to help themselves.

Cluster 4 includes activities which help the nurse to manage the environment, cluster 2 includes coordinating and advocacy activities, and cluster 1 includes those activities which help the nurse to adapt care to the individual needs of the patient and to accommodate those needs.

Starting at cluster 3 moving clockwise around the map, clusters include a combination of affect, cognition, and action. Cluster 3 includes affective elements of caring, primarily. The clusters from 7 to 1 (moving clockwise), can be taught and learned and are the professionalized aspects of caring. Cluster 3 includes innate qualities that a nurse should have and which are less subject to being learned.

The nurses also saw the "nursing process" expressed in the map. The nursing process is one foundation of nursing theory which frames all nursing activities in a four-part process: assessment, planning, implementation, and evaluation. Each step is part of a continuous feedback process. Cluster 1 is seen as assessment and evaluation, clusters 2 and 4 as planning, and clusters 7 to 5 as implementation. Cluster 3 includes the basic elements of a caregiver which provide the necessary foundation for the nursing process.

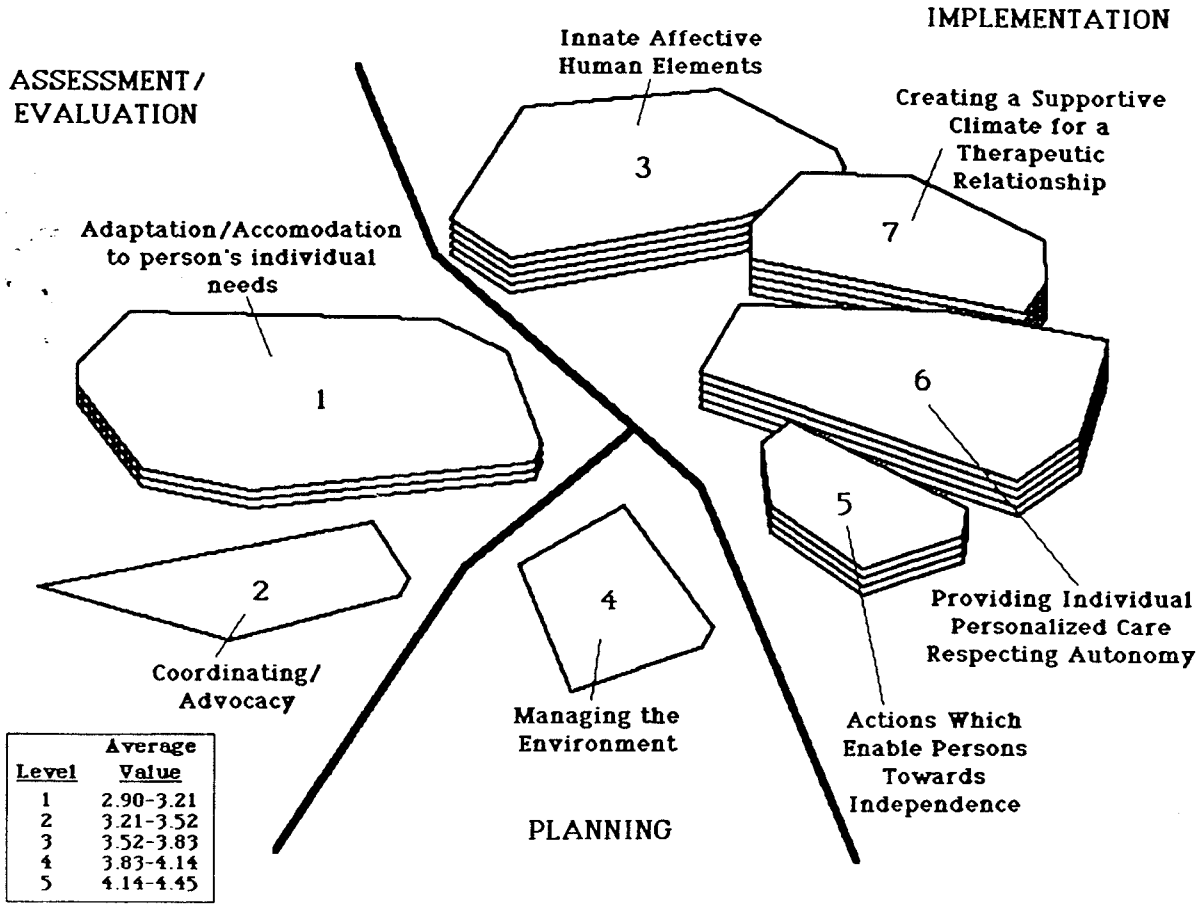


Figure 1. Nurses' concept map.

Patients. Figure 2 is the concept map for patients as they interpreted it.

Cluster 3 is a highly rated cluster which includes those integrated qualities of the nurse which increase trust between the nurse and patient and includes the way in which the nurse respects the patient's autonomy, and how honest and professional the nurse is with the patient. Clusters 6 and 7 are those clusters which demonstrate caring in action (based on the trust expressed in clusters 3 and 4). Cluster 6 is "informing the patient, and being informed." It involves the interactions between the nurse and the patient related to teaching and learning. Cluster 7 includes more activities which the nurse does for the patient to be "as helpful as possible." In addition, it includes the patient's family as recipients of these activities. Cluster 5 has less to do with individual patient/nurse interactions and more to do with formalized aspects of caring which involve caring for a group of patients. This includes maintenance of the environment or milieu.

Cluster 1 includes advocacy activities. Cluster 2 includes items which involve a higher degree of nurse control. They could be the paternalistic aspects of caring which foster dependency and include such things as "holding a patient's hand," and "doing what's best for the patient regardless if it is what's wanted." Patients describe these items as potentially non-caring items depending on the degree of control the nurse has versus the control the patient has. Patients want to have the control.

Theorists. Figure 3 shows the interpreted concept map for nurse theorists and researchers of caring.

Clusters 1, 2, and 3 are the basic elements of caring. Cluster 1 includes those elements associated with Watson's theory of care (1979) which emphasize existential, phenomenological processes between care giver and receiver. These terms are expressed in academic terminology. Clusters 2 and 3 are Leininger's basic elements of caring (1981, 1984) which are expressed in more common universal terminology. Clusters 5, 4, and 7 highlight the implementation of caring. Cluster 5 describes the teaching and learning process. Cluster 4 suggests items related to the coordination and maintenance of the milieu. Clusters 6 and 7 involve observable nursing behaviors which put caring into action.

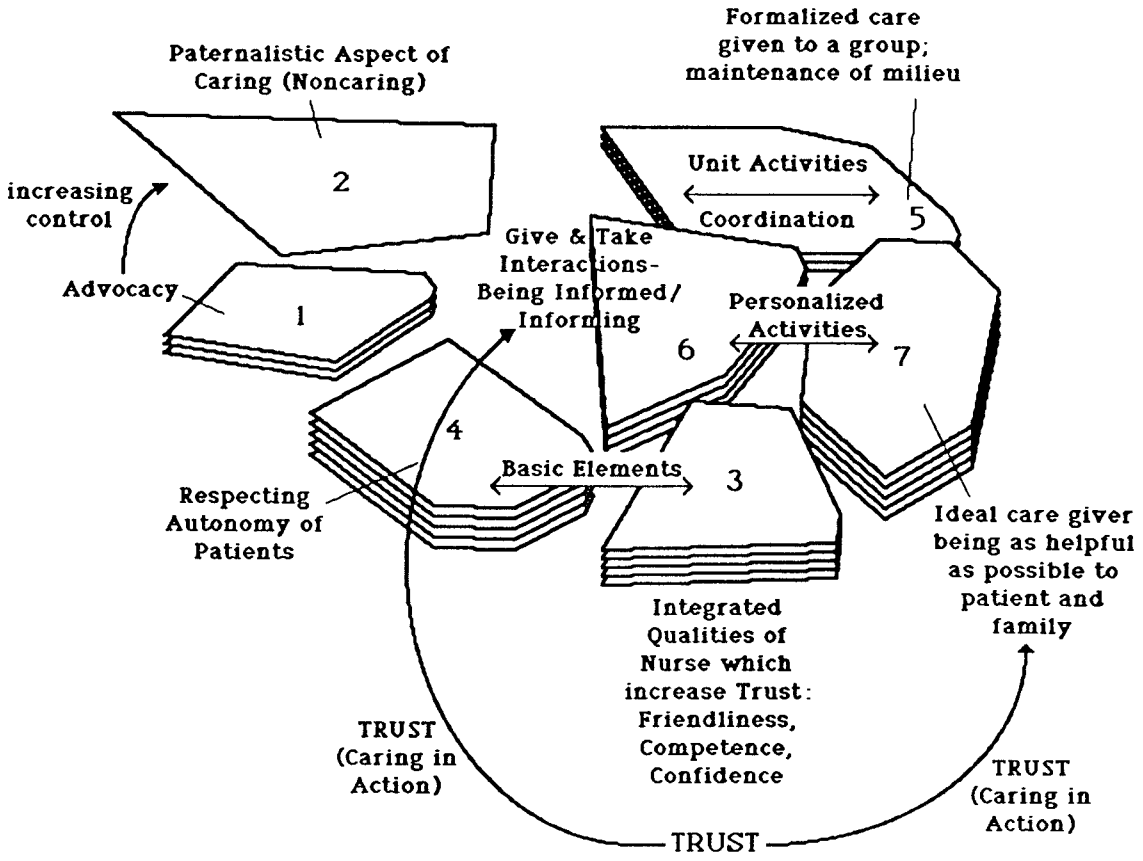


Figure 2. Patients' concept map.

Themes Common to All Three Maps

All three maps have some cluster of basic elements of caring, which are in close proximity to a cluster(s) highlighting interactions between a nurse and patient on a personal one-to-one basis. Each of the maps has elements of advocacy, and coordination and maintenance of the milieu. These are further away and less highly rated than the direct care giving (implementation) clusters described above. Each of the maps includes a dynamic element of putting caring into action. It is not enough to have caring feelings, they must be put into action which is often observable. Any questionnaire items which are developed should cover these basic areas of the conceptualization of caring; core elements, personalized care, maintenance and coordination of the milieu, advocacy, and action.

It is useful to look at the degree of correspondence between the maps. For instance, we can compare between the patient's cluster 3, the nurse's cluster 3, and the theorist's cluster 2. Each of these clusters represents the basic elements of caring. Items common to all three are warmth, being nonjudgmental (acceptance), kindness, trust, and compassion. Common between nurses and theorists were the items of touching, humor, and having a genuine attitude (authenticity). Common between the nurse and the patient were being honest; friendly; courteous and polite; and having confidence. Common between the patients and the nurse theorists was the item, listening (allowing people to verbalize).

It is also useful to examine the similarity in the clustering of items as well as the different clusters themselves. For example, in their main cluster, patients included such items as "the nurse being knowledgeable about an illness," "maintaining confidentiality," "giving patients choices," and "making the patient feel secure." These items were important aspects of the basic elements of care for patients. Therefore, any questionnaire would have to include these concepts when measuring patients' perceptions of caring. These same ideas were part of the conceptualization of caring for nurses; however, they were included in different clusters. A more detailed analysis of clusters reveals such differences as the one related to touch. Both nurses and theorists included touch in their most important central cluster of caring attributes, and both rated it fairly high (nurses, 3.60, theorists 3.82). Yet, patients included this item in their lowest rated cluster furthest away from the patients' central concepts of caring. Patients rated this item a 3.10, and it was in close proximity to item "hold their hand" which was rated a 2.40.

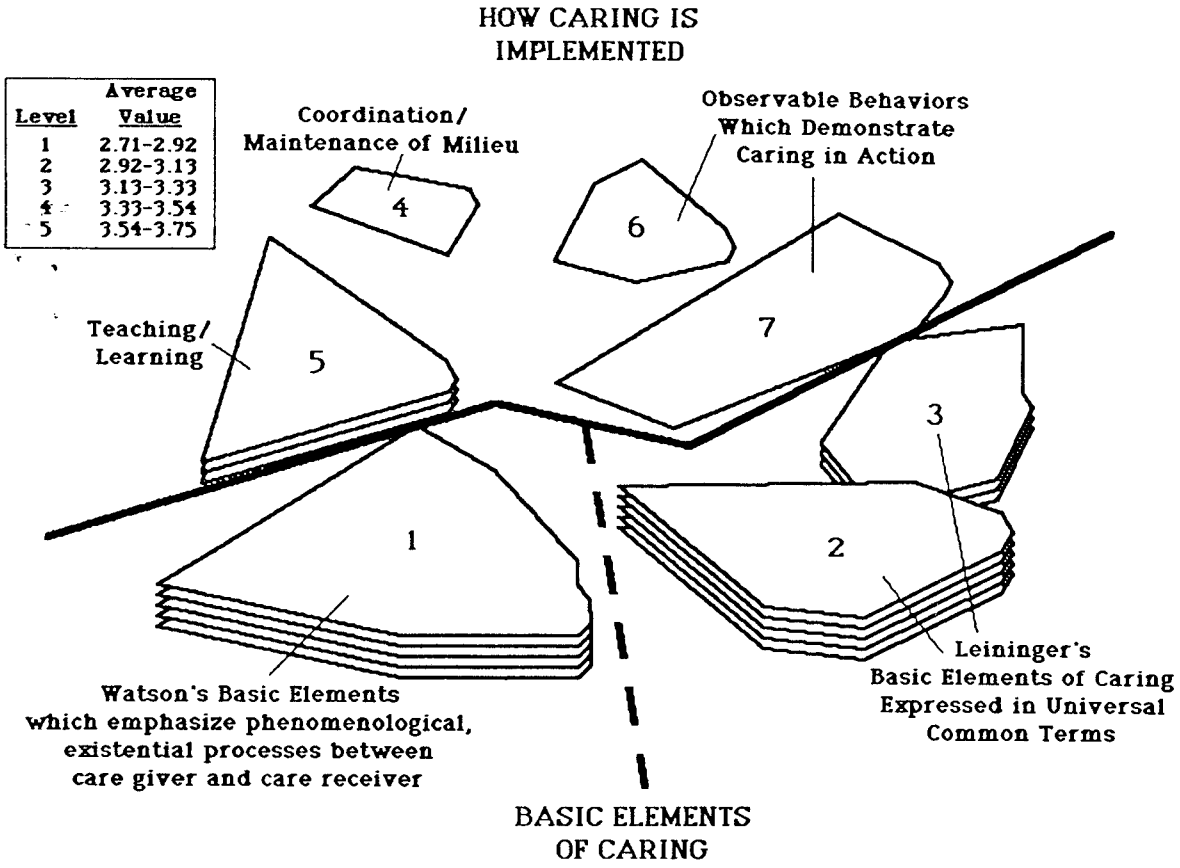


Figure 3. Nurse theorists' concept map.

Thus, there seem to be some differences in the way that patients, nurses, and nurse theorists think about touch. As patients discussed the map's meaning, they suggested that those items reflected a dependent and powerless position for the patient, a position which they rejected. Thus any measures of caring might reflect that patients will not value receiving touch, or having their hand held, as much as the nurses will value providing that aspect of caring.

CONCLUSION

One benefit in the use of structured conceptualization is that it provides a mechanism for engaging people in the explication of program theory (grounded theory) which can then serve as a theoretical pattern for construct validity when measures of a concept are obtained. In this case the concept maps for caring provided by the different participant groups can serve as the basis for predicting patterns in the obtained data. In addition, the process helps to define the conceptual domain so that it can be adequately represented in the measurement instrument.

Aside from the utilitarian benefits of using the structured conceptualization process, the technique provides a vehicle for group discussion of abstract concepts of importance to participants. The visual representation allows participants to see patterns and relationships which they may have intuitively felt were present but which they had never articulated. Participants often express that it is "fun" to engage in these activities. This feeling of goodwill helps to set the stage for cooperation and participant interest in other stages of a research project. It may even positively influence utilization of results. It is an energizing process which allows people to generate their ideas, see them displayed and then discuss their meaning. Structured conceptualization can help to make participation in research a more engaging and less formidable phenomena for program staff and clients.

REFERENCES

LEININGER, M. (1981). *Caring, an essential human need: Proceedings of the three National Caring Conferences*. Thorofare, NJ: Charles B. Slack.

LEININGER, M. (1984). *Care: The essence of nursing and health*. Thorofare, NJ: Charles B. Slack.

TROCHIM, W. (1985). Pattern matching, validity, and conceptualization in program evaluation. *Evaluation Review, 9*, 575–604.

TROCHIM, W., & LINTON, R. (1986). Conceptualization for planning and evaluation. *Evaluation and Program Planning, 9*, 289–308.

WATSON, J. (1979). *Nursing: The philosophy and science of caring*. Boston: Little, Brown and Company.

WATSON, J. (1985). *Nursing: Human science and human care—A theory of nursing*. Norwalk, CT: Appleton-Century Crofts.

Evaluation and Program Planning, Vol. 12, pp. 25-29, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

CONCEPTUALIZING FEMINISM

Clarifying Social Science Concepts

RHODA LINTON

Union Graduate School, The Union of Experimenting Colleges and Universities

ABSTRACT

This paper describes the use of structured conceptualization to develop a framework for understanding the concept of feminism. Ideas were brainstormed by 84 self-defined feminists and 150 of these ideas were randomly selected from the 710 which were contributed. Two other groups of self-defined feminists (total n = 34) conducted an unstructured sort of the ideas and the sort information was analyzed using multidimensional scaling and cluster analysis to produce a concept map. The map which resulted is described and some of the advantages of using this method are considered.

Feminism in its theoretical formulation is so broad, varied and conflicting that it escapes definition in specific terms. Attempts made over the past 10 to 15 years to capture and categorize its meaning and characteristics with various labels (e.g., Eisenstein, 1983; Freeman, 1975; Fritz, 1979; Jaggar, 1983) have contributed to a flourishing literature which continues to reflect feminism's diverse and rather elusive theoretical meaning. Although process is a definite focus for feminist theorists, not as much attention has been given to systematic recording, analyzing, debating, and synthesizing the meaning of feminism as it is embodied in practice. For example, what does the practice at the Seneca Women's Encampment for a Future of Peace and Justice tell us about the meaning of this concept (Linton, in press)? What is revealed about the meaning of feminism in the continuing struggle for reproductive rights, e.g., from legal and legislative campaigns to preparations to resist the growing number of increasingly violent attacks on birth control and abortion clinics?

There is often confusion about the relation of the terms "feminist" and "women." Some women see themselves as advocates of feminism;[1] others do not (Rowland, 1979). Advocates of feminism consider all women's experience as a basis for defining issues, although our own life experience and circumstances constrain what we know about "all women's experience." Further, many women who do not consider themselves advocates of feminism still support the efforts of advocates of feminism on many issues which benefit them specifically or which benefit women in general. They also, both individually and collectively, work on varying issues as women. Although this study concentrates on activities and beliefs of women who are advocates of feminism, it should be understood that the base of experience which constitutes the domain of our activities is the life experience of all women (for further examples, see references: Association of African Women for Research and Development, 1980; Beneria & Roldan, 1986; Davies, 1983; Kishwar & Vanita, 1984; and Russell, 1976).

Clearly the conceptual domain of this topic is rich in its variety of both substance and method. Accessing this domain in a way consistent with feminism's tenets and producing an outcome which is useful in the continuing clarification of the concept is the challenge facing our conceptualization method. The study I will cite to illustrate this method (Linton, 1985) had a dual purpose: to produce a conceptual map of the meaning of feminism to the particular participants involved; and to

[1]This term, advocates of feminism, is used by bell hooks (1984) in her recent book to indicate a belief in ending sexist oppression, rather than a belief in seeking sexual equality. I use it consciously to reflect her meaning.

Requests for reprints should be sent to Rhoda Linton, 150 Porter Hill Road, Newfield, NY 14867.

explore the fit of the conceptualization method to feminist principles as derived from feminist practice. This paper focuses on the former—the concept mapping process itself. In addition to raising questions about and clarifying our present knowledge and activity, the map can further provide a basis for thinking about and debating future theoretical and programmatic directions.

THE CONCEPT MAPPING PROCESS

In the implementation of the study, a group-centered approach to the three steps of the general model (i.e., generation, structuring, and representation) was followed. I refer to these steps in a perhaps more common language as expansion, contraction and interpretation. Ideally, all three steps are done with the same group in the context of working toward accomplishing some programmatic goal. However, in this study, two different groups contributed to steps 1 and 2, and participant voices from both steps 1 and 2 constitute the basis of the interpretation step (3). (It should be noted that this study was limited to participants in the U.S.)

Expansion

The ideas for the conceptual domain were contributed by 84 (of 205 requested) self-defined advocates of feminism from 21 states who were randomly selected from one predominantly white women's group and two national level gatherings addressing the issues of women of color and third world women's perspectives on feminism (Linton, 1985, pp. 51–80 give a detailed description of the participants in both steps 1 and 2). In a mailed survey participants were asked to "brainstorm feminism." Of the 710 ideas contributed, 150 were randomly selected to constitute the conceptual domain.

Contraction

The participants in this step were 34 self-defined advocates of feminism who attended meetings to organize the ideas; one meeting was held in Ithaca, New York (26 attending) and one, a week later, in New York City (8 attending). Each participant received a stack of 150 cards containing the ideas selected and was asked to sort them into piles which made sense to them. Recording systems were provided. The data was then analyzed using multidimensional scaling and cluster analysis procedures. Graphic representations were in the form of a two-dimensional map of the ideas with clusters identified by symbols. An additional task for participants was to rank each idea on a scale of from 1 (least) to 5 (most) on its level of importance to feminism.

Interpretation

In an ideal situation, participants meet in a group and are provided with copies of the cluster map, lists of the ideas by cluster with their ID number, and their average level of importance to feminism value. They then engage in a process of negotiation to come to agreement on naming the clusters based on their thematic content, and to possibly name regions of the map based on cluster themes and location. Further exploration of the meaning of the conceptual framework utilizes the familiar conceptual tools of distance and direction to understand the meaning of the location of the ideas and their interrelationships. The level of importance to feminism indicator can be used as a height dimension so that "mountainous" areas of the map can be interpreted as more important.

In my non-ideal study, the interpretation of the map was limited by my lack of access to the participants. Given this condition, the interpretation of the meaning of the map is limited. However, some impressions can be given to indicate some possible meanings.

First, an example of a general meaning can be seen in the names of the clusters as shown on Figure 1. The names are based on the similarity in meaning of the ideas which compose them, and give an idea of the breadth of the thinking about the concept. There is no inherent order of importance in their listing.

Second, the general meaning of the concept map can be approached by viewing the map as a whole. When doing so, every entity is represented in relation to every other entity. The organization of the entities into clusters and the arrangement of these clusters and the regions which they form provide a constant and undeniable picture of the interrelationship among all of the individual ideas. Viewing the map as a whole can reveal direct as well as underlying connections. For example, there is clearly a direct connection between both the meaning and the location of Clusters 7 and 8, creating a region of clusters relating to feminist theory. Another example of direct connections can be seen among Clusters 2, 4, and 9, a region of those clusters relating to specific and practical goals, actions, and changes. The relationship between these two regions, which lie directly across the map from each other, is an example of an underlying connection. This particular underlying connection could be viewed as a dimension connecting theory and practice. Comparisons of the content of these clusters could reveal whether what advocates of feminism actually do reflects what we say we are doing—and vice versa. For example, one entity from Cluster 8 is "feminism concerns itself with a broad spectrum of issues of concern to women." Such issues are specifically noted in Cluster 9: "quality, affordable housing;" "an absolute stop to involuntary sterilization;" "cooperative business enterprises and housing;" "no more torture;" and "complete overhaul of welfare system so that women aren't trapped in per-

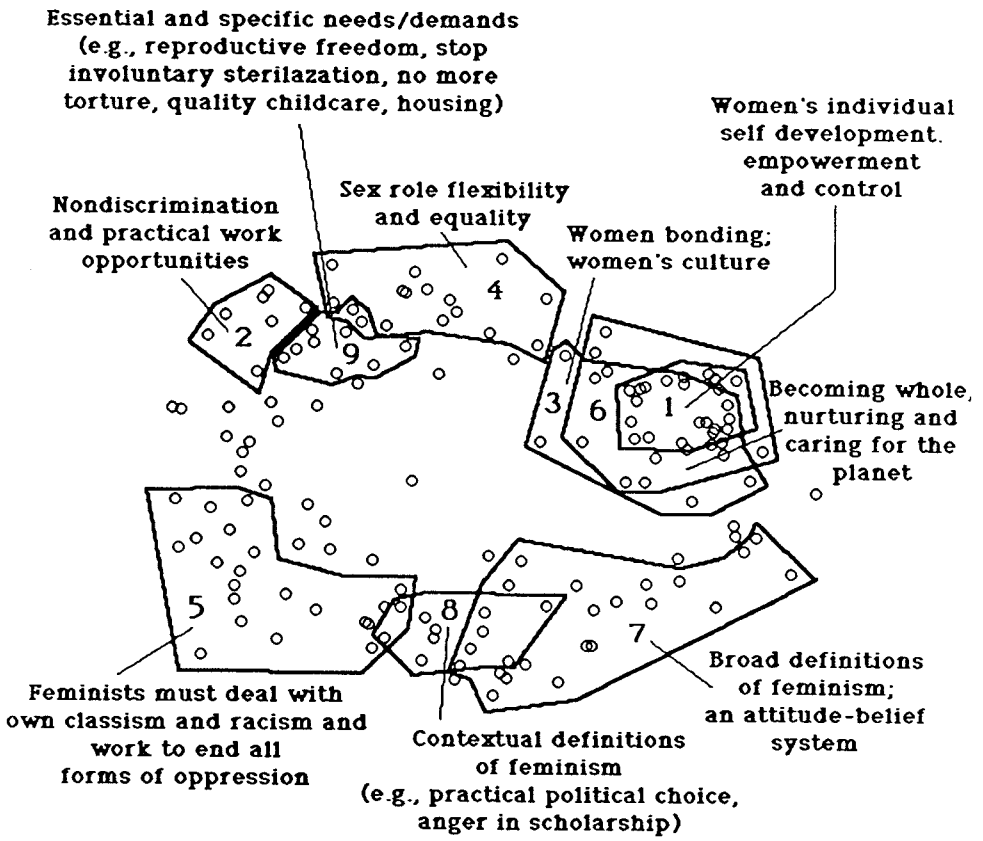


Figure 1. Concept map of "feminism."

petual dependency on the state." Underlying connections can also reveal inconsistencies in our thinking. For example, consider the underlying connection between Clusters 1 and 5. The dimension connecting these two clusters could be interpreted in several ways: it could be called self and other, or personal and political, or domestic and global, all very familiar ideas developed in both theory and practice over the past two decades. There is one contradiction to this connection, however. Entities claiming our needs as advocates of feminism to deal with our own classism and racism fall in the "other oppressions" cluster (5) rather than into the personal/self development and change cluster (1). There apparently is a difference in how we think about personal changes relating to other oppressed groups and personal changes relating to our own self development, affirmation, and empowerment (i.e., personal "growth"). Confronting our own racism and classism somehow is viewed as being in a cluster of "other" rather than "self." How and why this occurs, and what we do about it, are only some of the questions raised by this information. Viewing the map as a whole can also provide some insight into possible action on such questions. For example, it can be proposed that movement from Cluster 1 to Cluster 5 requires either a huge leap across the "unknown center" or going by way of theory in one direction or practice in the other. In other words, in order for advocates of feminism to connect our selves and personal development to dealing with our own classism and racism, as well as working to end all oppressions, we have to go through some theoretical or practical experience. It doesn't just happen. The specific ideas in either of those regions may give direction as to how and/or why this particular separation exists, and how to move to change this condition of separation.

One further general observation about the cluster relationships, when viewing the map as a whole, is that when moving from the top of the map to the bottom (i.e., along the vertical axis), cluster meanings change from specific to general. Similarly, when moving from side to side (i.e., along the horizontal axis), from west to east, cluster meanings change from other, or global, to individual, or self. This information forms an overall framework within which to view the individual ideas, clusters, regions, and their interrelationships.

Third, an indication of what the group considered important can be seen by individual idea and cluster level of importance to feminism values. Values closer to 5 are higher. Two ideas had values between 4.6 and 5.0

and could be considered the two most important ideas in the study. The highest value (4.6667) was given to the idea "recognizing the interconnectedness of the struggles of all oppressed persons—working to end all oppressions" in Cluster 5, and the second highest value (4.6061) was given to the idea "control of own life" in Cluster 1. The idea with the lowest value (1.2308) was "being more feminine" in Cluster 1. Cluster averages reveal that Clusters 2 (Nondiscrimination and practical work opportunities) and 3 (Women bonding, women's culture) were considered most important, and Clusters 7 (Broad definitions of feminism, an attitude-belief system) and 8 (Contextual definitions of feminism, e.g., practical political choice, anger in scholarship) least important. When considered in the context of the other clusters in the two regions represented by clusters 2 and 8, these level of importance values raise interesting questions. Clusters 2 and 9 contain action oriented entities relating to specific practical needs/demand including an emphasis on work opportunities. Cluster 3 includes an emphasis on women actively relating to each other by exchanging support and joining together to strive for common goals. Clusters 8, 7, and 6 contain more passive entities, such as ideas of what feminism is about, and descriptions of the nurturant and caring, holistic views of women. This comparison indicates that advocates of feminism appear to value action on specific issues more than thinking about them and, perhaps, more than thinking about the meaning of those actions. This could mean that in our commitment to change what has been labeled our passivity and resulting victimization, we could be acting on a short term, short-sighted basis in ways that may conflict with our long term, more globally stated beliefs.

CONCLUSIONS

On the basis of this study some impressions of the advantages of using the concept mapping process in this context for clarifying the concept of feminism for both theoretical and programmatic purposes are in order. It is clear that this method of conceptualization is an open, moving, dynamic process which is responsive to rather than manipulative of participants; which includes data, the format of which allows for inclusion of a wide diversity of similar and conflicting ideas; and which is accessible to interpretation by a wide population. It can be used within groups and across groups, producing conceptual maps for comparison and as a basis for single or joint group discussion, evaluation, and planning.

Further, the process itself tends to defuse potentially hostile and defensive feelings and situations, and to support the growth of group unity. Some reasons for this are: there is equal power and influence of all participants in all phases of the process (with the possible exception of group brainstorming where power relations may be unbalanced); the agreed-upon final set of contributed ideas is included in the final conceptual map; there is easy accessibility to understanding the meaning of the map; there is a lack of pressure attached to the process because its orientation is to how specific parts relate to the whole rather than isolating them to sink or swim on their own; the group synergism creates a whole which is obviously more than the sum of its parts, validates the individual's investment of trust, and energizes rather than drains the group; the interpretation of the map places what is already known into a configuration which incorporates new information in both substance and method.

The following list provides a short summary characterizing advantages of this method:

- it invites many and diverse voices to be heard
- voices can be gathered which have not had access to being heard in print
- it asks for women's own words to name our reality
- it is grounded completely in women's life experience and thought
- it is non-judgmental, accepting all contributions
- all participants' contributions have equal power
- the group setting for structuring works to balance power between participants and researcher
- it requires a minimum of intrusion of the researcher's opinion in all three steps
- it does not eliminate any of the ideas; they all appear on the map
- it presents all of the ideas in relation to each other
- it reveals simultaneously each entity in relation to all others and to its own cluster-mates, cluster relations overall and regionally, as well as the total concept, all in a holistic view
- it emphasizes connections rather than "significant' differences
- it actively involves the participants in the knowledge creation process
- the process can be as important as the product, esp cially when done in a single group
- it provides a possible learning experience for pa ticipants in exchange for helping the researcher
- it tends to create a positive feeling among parti pants, especially when all three steps are done wi an on-going group

- it tends to create a feeling of ownership of the product among participants
- it results in the researcher not being totally isolated
- throughout the steps there is a strong collaborative, cooperative aspect among participants and including the researcher
- the map provides open access to interpretation because it is a widely familiar format
- it does not exclude people from understanding its results if they do not have knowledge of the use and meaning of special research techniques such as statistics and graphics

However, although western advocates of feminism may be ready to utilize this method, recent discussion with advocates of feminism working in other countries indicate less potential. For example, women planning an international youth conference rejected the method based on its use of western high technological machinery which they distrust as a weapon of imperialism. Caribbean advocates of feminism in highly educated circles like the method but don't see how it can be used with non-literate women, while an English researcher, who trains women from the third world in research techniques, thought it would be useful in understanding some crucial problems occurring in women's cooperatives, but questioned the women's access to computers in their own countries, as well as their suspicions of the priority given to western high tech methods in the face of the low priority given to solve more fundamental life problems such as famine, starvation and the imposition of export economies in third world countries. These comments clarify one point where the method is clearly not consistent with feminist activity/principles; that is, although the participants have major control of the conceptualization meaning, and decision making about much of its structure and process, they are dependent on the researcher for the technology. Although there are ways to minimize this dependency in some situations (e.g., by the use of skill exchange in the study), it is a potentially serious limitation to use of the method to clarify the concept of feminist theory and practice from a more global perspective.

REFERENCES

ASSOCIATION OF AFRICAN WOMEN FOR RESEARCH AND DEVELOPMENT. (1980). Genital Mutilation: A Statement by the Association of African Women for Research and Development (AAWORD). *Resources for Feminist Research, 9* (1).

BENERIA, L., & ROLDAN, M. (1986). *The crossroads of class and gender*. Chicago, IL: University of Chicago Press.

DAVIES, M. (1983). *Third World/Second sex, women's struggles and national liberation*. London: Zed Books Ltd.

EISENSTEIN, H. (1983). *Contemporary feminist thought*. Boston: G.K. Hall & Co.

FREEMAN, J. (1975). *The politics of women's liberation: A case study of an emerging social movement and its relation to the policy process*. New York: David McKay.

FRITZ, L. (1979). *Dreamers and dealers: An intimate appraisal of the women's movement*. Boston: Beacon Press.

hooks, b. (1984). *Feminist theory, from margin to center*. Boston: South End Press.

JAGGAR, A. (1983). *Feminist politics and human nature*. Totowa, NJ: Rowman and Allanheld.

KISHWAR, M., & VANITA, R. (1984). *In search of answers*. London: Zed Books, Ltd.

LINTON, R. (1985, March). *Conceptualizing feminism: A structured method*. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

LINTON, R. (in press). Seneca Women's Peace Camp: Shapes of things to come. In A. Harris & Y. King (Eds.), *Rocking the ship of state: Toward a women's peace politics*. Boulder, CO: Westview Press.

ROWLAND, R. (1979). *Women who do & women who don't join the women's movement*. London: Routledge & Kegan Paul.

RUSSELL, D.E.H., & van de VEN, N. (Eds.). (1976). *The Proceedings of the International Tribunal on Crimes Against Women*. East Palo Alto, CA: Frog In The Well.

Evaluation and Program Planning, Vol. 12, pp. 31-36, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

CONSTRUCT VALIDITY IN MEASUREMENT

A Pattern Matching Approach

JAMES E. DAVIS

University of Delaware

ABSTRACT

*Construct validity refers to how well operationalizations in research reflect the theoretical constructs they are supposed to reflect. The multitrait-multimethod (MTMM) matrix framework proposed by Campbell and Fiske (*Psychology Bulletin, 56, *81-105, 1959) is the most widely known and used method for establishing construct validity. In determining if there is construct validity using the MTMM approach, researchers should have some idea or expectation of the interrelationships among constructs and the methods used to measure them. Structured conceptualization can be used to articulate the expected relationships between constructs that are being measured. The resulting theoretical concept map can then be compared with correlations in the MTMM matrix using a pattern matching approach to assess the degree of construct validity evidenced.*

Construct validity is the primary focus of this article. The procedure for establishing construct validity in measurement using a "pattern matching" approach is explored. In this approach, conceptualized theoretical patterns of construct interrelationships are matched with obtained patterns of interrelation among constructs. Advantages of using pattern matching for establishing construct validity are outlined. Also, several different analyses for indicating a match between expected and obtained patterns are presented.

CONSTRUCT VALIDITY

A major conceptual view of construct validity is the nomological network presented by Cronbach and Meehl (1955). This nomological network assumes that there is an interlocking system of principles which constitutes a theory and establishes the relationship between theoretical constructs and observed measurements. The nomological net is an attempt to articulate and clarify the concept of construct validity. It requires that one explain relevant constructs and set them in a relational scheme with observed measures and other constructs. Within this framework, explicit statements about constructs are required in order to claim construct validity. Campbell and Fiske's (1959) multitrait-multimethod matrix (MTMM) framework approach is an example of making explicit statements of relationships about constructs and observed measures and this can be viewed as one way to approach the development of the nomological net.

Multitrait-Multimethod Matrix (MTMM)

The underlying assumption of the MTMM is that different methods are being used to measure each construct. Within this framework, constructs are referred to as traits. A trait is a continuous latent variable that is inferred from observable data. The objective of a measure is to tap traits of interest, such as achievement, personality, or sociability. Method may be defined as a multiple-choice test, open-ended answer test, questionnaire, face-to-face interview, or unobtrusive data collection technique. Different rating scales may be viewed as different methods. Also, different item types of a multiple-choice instrument, such as analogy items

Requests for reprints should be sent to James Davis, Department of Educational Studies, Hall Education Building, University of Delaware, Newark, DE

or analytic reasoning items can be considered examples of different methods. Hence, the idea of method encompasses measurement instruments as well as variations within measures. The assumption that one method only reveals a part of a social or a psychological reality is a prevalent one. It is generally felt that a single measuring instrument is insufficient to comprehensively measure a construct. Furthermore, with only one method of measurement, a researcher has no way of distinguishing trait variance from unwanted method variance (Campbell & Fiske, 1959).

Trait variance is the variation within a measure which is attributable to the true value for that trait for a set of individuals. Method variance refers to variation in measures that is attributable to the effect of the method on the observed score. Trait variance is of primary concern in establishing construct validity, since a given measure should tap the theoretical construct without interference from method of measurement. What distinguishes method from trait is that the systematic variation due to method is undesirable. A researcher wishes to measure traits and not methods.

The rationale for using the multitrait-multimethod approach for construct validity is that if methods are independent and traits are accurate in their assessment, then the relationship among traits should stay the same across different methods. Also, any error due to method should be equally observable with each trait that is measured by that method. The Campbell and Fiske MTMM matrix describes a classification system where each measure represents a trait-method combination of one trait with one method. Campbell and Fiske make a distinction between two types of construct validity; convergent and discriminant. Evidence for these two types of construct validity is based on interpretation of the MTMM matrix. To understand the criteria on which the interpretation is based, an MTMM which has the same structure as the one discussed by Campbell and Fiske (1959) is presented in Figure 1 and represents a hypothetical case where there are three methods crossed with three traits. First, there must be convergence, or agreement among indicators which claim to measure the same construct. This is expressed by high correlations among assessments of the same trait with different methods as shown by the convergence diagonal illustrated by the C's in Figure 1. Second, there must be a divergence between measures of similar but conceptually different constructs. So, a researcher would expect a low correlation of measures of different traits. The same method triangles (M) and the heterotrait-heteromethod triangles (H) show this pattern. By having convergence among measures of the same trait and divergence of measures for different traits, the case for construct validity is supported. In general, it is expected that the reliability coefficients (R) are larger than the convergence diagonals (C). Also, the convergence diagonals are expected to be higher than values in same method triangles (M), while the (M) triangles are expected to have larger correlation coefficients than the triangles representing different traits measured with different methods (H). In order to assure the measurement of a trait with independent methods, one has to be able to trace theoretical patterns in measurement across different methods. Similarly, a method effect can be established as a pattern of measurement consistent across traits but not across methods (Wothke, 1984). In the MTMM framework, the researcher must articulate some expected patterns of interrelationships between traits and methods to compare with relational patterns that are characteristic of the MTMM. This comparison should reveal the similarity of expected patterns and the patterns underlying the MTMM.

		METHOD 1			METHOD 2			METHOD 3		
	TRAITS	A	B	C	A	B	C	A	B	C
1	A	R								
	B	M	R							
	C	M	M	R						
2	A	C	H	H	R					
	B	H	C	H	M	R				
	C	H	H	C	M	M	R			
3	A	C	H	H	C	H	H	R		
	B	H	C	H	H	C	H	M	R	
	C	H	H	C	H	H	C	M	M	R

R - Reliability of method

C - Covergence diagonal

M - Same Method triangles

H - Heterotrait-Heteromethod triangles

Figure 1. Structure of a hypothetical multitrait-multimethod matrix with three traits and three methods.

In summary, the notion of construct validity as conceived by Cronbach and Meehl cannot be expressed in just a single coefficient. The establishment of construct validity requires several pieces of evidence, not all of them quantitative. Campbell and Fiske's MTMM matrix is the most widely recommended method for establishing the construct validity of measurement. Simple rationalizing or theorizing about the nature of a construct and its relations to other constructs is not sufficient for establishing construct validation. Measures which are taken to be observable expressions of a construct are expected to yield data that are consistent with how one expects constructs to be interrelated. Campbell and Fiske's MTMM was developed based on these requirements and continues to be the most widely recommended method for establishing the construct validity of measurement.

A pattern matching approach for construct validity

has been proposed by Trochim (in press). As previously mentioned, researchers should have some idea or expectation of how constructs are interrelated. Trochim suggests the use of a structured conceptualization method that is helpful in articulating the expected relationship between constructs that are to be measured. This method focuses on theorized expected patterns of measurement interrelationships and how they are verified using actual obtained data. Generally, the primary task in this approach is to match patterns of theoretical constructs' interrelationships with operationalized obtained data patterns. The next section describes a hypothetical example of pattern matching and how it may be used to establish construct validity in measurement.

PROCEDURE

Data

In this example, data from Campbell and Fiske's (1959) synthetic multitrait-multimethod matrix serve as the operationalized obtained pattern of interrelationships among measures (see Figure 2). These data adhere to the multitrait-multimethod criteria that were previously described. The Campbell and Fiske matrix is then compared to four theoretical matrices which also meet the multitrait-multimethod criteria and which could be generated by a variation of the structured conceptualization approach. By differing the degree to which method and trait relationships are specified, each of the four generated matrices represents a different theoretical pattern of measurement interrelationships. A hypothesis is made that the more theoretical information a researcher is willing to specify about the interrelationship of methods and traits, the better the researcher is able to demonstrate construct validity. In this case, for instance, it is expected that theoretical matrices which specify more information about method and trait patterns of interrelationship will more closely approximate the obtained Campbell and Fiske matrix. The assumption is made that a better match with the obtained correlation matrix pattern indicates greater construct validity of measures.

To summarize, the interrelationship of measures in the Campbell and Fiske synthetic matrix represents the obtained correlational pattern, while the four other matrices represent different theoretical specification of the patterns of relationships.

Deriving Theoretical Patterns

A method matrix and a trait matrix are used to specify the four theoretical patterns (see Figure 3). Three methods and three traits are assigned to each matrix corresponding to the structure of the Campbell and Fiske matrix. The choice of three methods and three traits is more a function of convenience and attempting to duplicate the Campbell and Fiske matrix than of some theoretical consideration. We can develop the theoretical MTMM matrices by specifying the expected method and trait relationship separately in these four 3×3 matrices. For instance, if we believe that traits are all equally interrelated, we would construct a matrix which reflects this. In Figure 3, the 2's along the method main diagonal are noticeably smaller than the 4's along the trait matrix main diagonal. This difference is not a crucial one. It is done in order to satisfy a requirement for construct validity, the trait variance is dominant over method variance. The selection of specific diagonal values of the four theoretical patterns is in part to avoid

		METHOD 1			METHOD 2			METHOD 3		
	TRAITS	A	B	C	A	B	C	A	B	C
1	A	100								
	B	58	100							
	C	38	37	100						
2	A	57	22	9	100					
	B	22	57	10	68	100				
	C	11	11	46	59	58	100			
3	A	56	22	11	67	42	33	100		
	B	23	58	12	43	66	34	67	100	
	C	11	11	45	34	32	58	58	60	100

Figure 2. Campbell and Fiske's (1959) hypothetical multitrait-multimethod matrix.

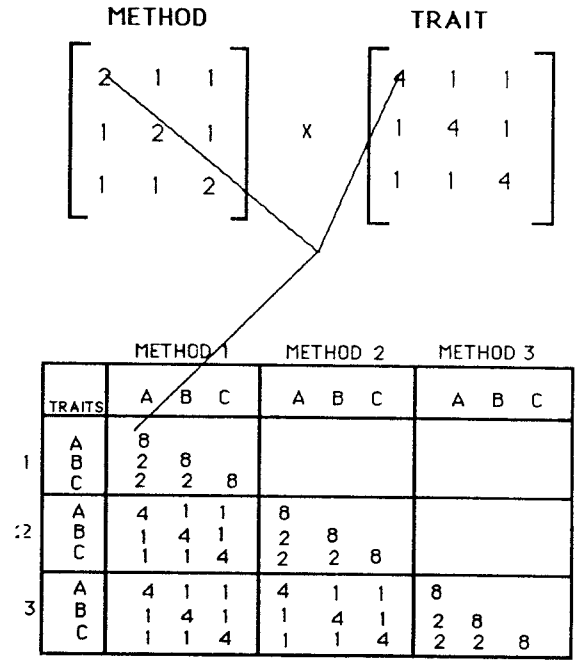


Figure 3. Procedure for deriving theoretical patterns by cross-multiplying trait and method matrices.

degenerate multidimensional scaling solutions (Davison, 1983). The theoretical patterns of interrelationship are indicated in the off diagonal values. In Figure 3, the assumption is made that methods and traits are equally interrelated. To obtain the full MTMM for a theoretical specification, the theoretical method and trait patterns are multiplied to produce an MTMM that meets the Campbell and Fiske criteria for establishing construct validity.

As previously mentioned, the four theoretically derived matrices represent varying degrees of specification of the method and trait interrelationship as illustrated by Figure 4. In the first pattern, methods and traits are theorized to be equally interrelated. The second pattern illustrates methods as not being equally interrelated while traits are. Traits are not assumed to be equally interrelated but methods are in the third pattern. Finally, a theoretical assumption is made that both methods and traits are not equally interrelated in the fourth example. In all cases when interrelationships are not equal, values for the matrices are based on the apparent patterns in the Campbell and Fiske MTMM. In practice, however, the researcher would specify expected relationships on the basis of theory.

Configural Similarity

The multitrait-multimethod matrices from the four different theoretical patterns are compared to the Campbell and Fiske matrix. The maps generated by multidimensional scaling of the theoretical matrices are then compared to a map generated by using the correlations from the obtained Campbell and Fiske synthetic matrix. First a multidimensional scaling analysis was conducted (Davison, 1983; Kruskal & Wish, 1978) on the multitrait-multimethod correlation matrix. The analysis renders a two dimensional map where every method and trait is represented in relation to every other method and trait as shown in Figure 5. The figure also shows the map derived for one of the four theoretical patterns described above. Here, the theoretical map (which is at the bottom of Figure 5) represents the example of assuming equal trait interrelationship only. As expected, the theoretical map illustrates pictorially how the traits and methods cluster together.

Each point on the map represents one of the nine method-trait combinations from the MTMM. In viewing the map in Figure 5, we can see the patterns of interrelationship among methods and traits. The Campbell and Fiske obtained map reveals distinct method

	Interrelationship Method	Trait	METHOD	TRAIT
1.	NO	NO	$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$	$\begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{bmatrix}$
2.	YES	NO	$\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
3.	NO	YES	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$
4.	YES	YES	$\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$	$\begin{bmatrix} 7 & 2 & 1 \\ 2 & 7 & 1 \\ 1 & 1 & 7 \end{bmatrix}$

Figure 4. Theoretical patterns for different specifications of trait and method interrelationships.

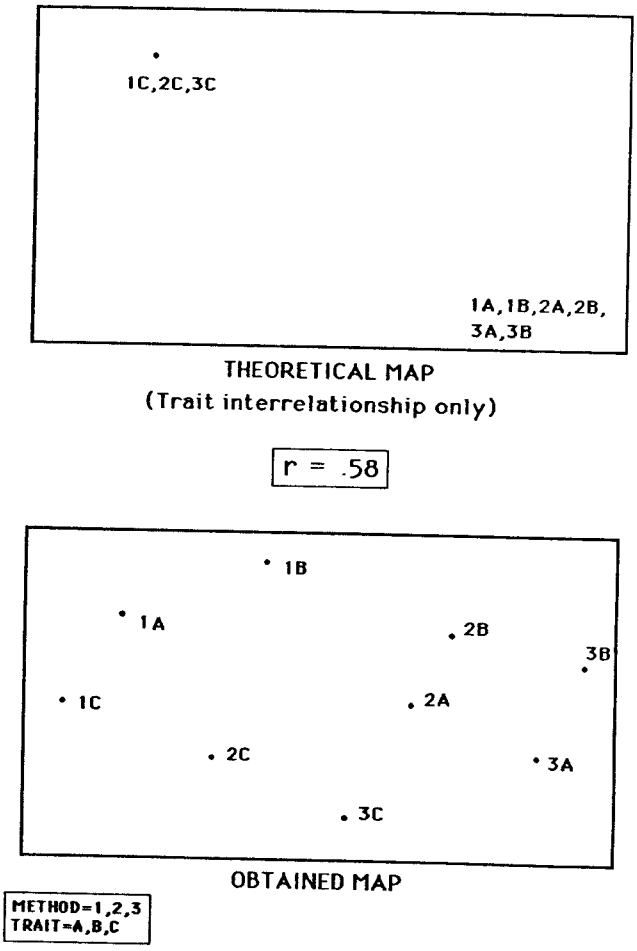


Figure 5. Pattern matching for construct validity where trait interrelationships alone are specified.

and trait clusters, while the theoretical map shows a separate clustering of the more dominant C-trait and a cluster of the A and B traits.

Figure 5 is an example of a theoretical map that reflects a theoretical pattern where methods are interrelated differently and traits are equally interrelated. The configural similarity between these two maps (i.e., the theoretical and obtained) is derived by using a Pearson's Product Moment Correlation of Euclidean distances between the points on each map (Davison, 1983; Guttman, 1968; Kruskal, 1964). This correlation coefficient is an indicator of how similar the two multidimensional maps are, and it indicates the match between the obtained and theoretical patterns of measure interrelationship. If the configuration of points on two maps are similar, the pattern of Euclidean distances will be similar, thereby increasing the correlation. While other methods of assessing configural similarity of multidimensional maps have been proposed (Gower, 1975; Schonemann & Carroll, 1970), the Pearson Product Moment Correlation appears to be a more familiar and understandable method. When using the Pearson correlation, higher coefficients indicate greater similarity between the maps. In other words, high coefficients are indicative of a pattern match and greater evidence for construct validity.

RESULTS

The four theoretical patterns which assumed different degrees of trait and method interrelationshp produced different correlation coefficients when matched with the Campbell and Fiske obtained matrix (see Figure 6).

The degree of similarity in the maps increased as more information about how traits and methods are related is taken into consideration. The first example where the least amount of information was assumed about method and trait interrelationships produced an r equal to .34. When methods were not assumed to be equally interrelated, the r increased to .54. Assuming differential interrelationship of traits produced an $r =$ .58. Finally, under the assumption that both methods and traits are not equally interrelated, the highest coefficient of .71 was obtained. By including more specific information about the interrelationships among traits and methods, the fourth theoretical pattern most closely approaches the pattern of the Campbell and Fiske correlational matrix. The higher correlation coefficient of the fourth pattern indicates greater similarity between the two maps. This evidence supports the assumption that when more information is specified about the theoretical expected patterns of interrelationships among traits and methods, one is able to obtain higher estimates of construct validity.

	INTERRELATIONSHIP		SIMILARITY
	METHOD	TRAIT	r
1.	NO	NO	.34
2.	YES	NO	.54
3.	NO	YES	.58
4.	YES	YES	.71

Figure 6. Similarity (pattern match) correlations between theoretical and obtained maps for four different trait and method specifications.

CONCLUSION

Pattern matching based on mapping techniques may be a useful approach for establishing construct validity. After theorizing the expected patterns of how traits and methods are interrelated, an attempt to verify these patterns using obtained data becomes the objective of pattern matching. This strategy offers several advantages. First, the logic of pattern matching as described here is consistent with the logic of the MTMM framework. Second, pattern maps represent the relationships among traits and methods, while providing an opportunity for exploring the effects of trait and method factors and interaction between them. Third, by utilizing multidimensional scaling one is able to take ordinal-level estimates of theoretical interrelationships and scale these into interval level theoretical patterns. A researcher might articulate pattern relationships by sorting or ranking or rating the measurement concepts and the multidimensional scaling analysis will scale these non-interval input data, thereby laying the foundations for correlations between patterns. Finally, pattern matching provides a single indicator (r) of construct validity. Prior to this, the interpretation of an MTMM was done simply by visually assessing the obtained correlation matrix.

Given these advantages, pattern matching which is based on structured conceptualization appears to be a viable approach for examining construct validity. Of course, this approach needs to be tested on a broader range of examples—both simulated and real—before we can recommend it with confidence. Nevertheless, if pattern matching proves to be a reliable method for assessing construct validity in measurement, the current strategies used by researchers to establish construct validity may be changed dramatically.

REFERENCES

CAMPBELL, D.T., & FISKE, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

CRONBACH, L.J., & MEEHL, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

DAVISON, M.L. (1983). *Multidimensional scaling*. New York: John Wiley and Sons.

GOWER, J.C. (1975). Generalized procrustes analysis. *Psychometrika, 40*(1), 33-50.

GUTTMAN, L.A. (1968). A general nonmetric technique for finding smallest coordinate space for a configuration of points. *Psychometrika, 33*, 469-506.

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

KRUSKAL, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*, 1-27.

SCHONEMANN, P., & CARROLL, R. (1970). Fitting one matrix to another under choice of a central dilation and rigid motion. *Psychometrika, 35*(2), 245-255.

TROCHIM, W.M.K. (in press). Pattern matching and program theory. *Evaluation and Program Planning*.

WOTHKE, W. (1984). *The estimation of trait and method components in multitrait-multimethod measurements*. Unpublished dissertation, The University of Chicago.

Evaluation and Program Planning, Vol. 12, pp. 37–43, 1989
Printed in the USA.

0149 7189/89 $3.00 + .00

A PATTERN MATCHING APPROACH TO ASSESS THE CONSTRUCT VALIDITY OF AN EVALUATION INSTRUMENT

JULES M. MARQUART

Cornell University

ABSTRACT

Pattern matching is an approach for determining the correspondence between program theory and the observations or data collected from a program based on that theory. For each major component of an evaluation—program, participants, measures, and outcomes—there are theoretical and observed patterns, and the degree of correspondence between the theoretical and observed patterns is called the pattern match. This chapter describes a measurement pattern match and its use to provide evidence for the construct validity of an instrument developed for a program evaluation. The evaluation context is a study of the effects of a child care program provided by a hospital on employees' work-related attitudes and behaviors. The intercorrelations from the questionnaire administered to employees provide the observed pattern. The structured conceptualization method was used to develop the theoretical pattern from the perspective of administrators of hospitals that provide child care assistance. Results are presented of both visual and quantitative assessments of the match between the theoretical and observed patterns.

Pattern matching is used here to mean the "fit" or correspondence between theory, however well articulated, and the observations or data collected to bear on that theory—that is, the "pattern matching of the conceptual and operational domains" (Trochim, 1985, p. 575). This approach builds on Campbell's (1966) idea that pattern matching between theory and data is a way of achieving scientific understanding. The value of the pattern match is that the validity of the conclusions drawn is strengthened if the pattern predicted by the theory is found in the data, because the likelihood that such a pattern of results could have occurred by chance is small.

Trochim (in press) has proposed a model based on pattern matching principles that assists the researcher in articulating theory and in integrating theory more directly into the research process. His pattern matching model is based on the interplay between the theoretical realm (which is conceptualized as the theoretical pattern of a program and includes theories, ideas and hunches about a program), and the observational realm (which is considered the observed pattern and is operationalized as the observations, measures, and data from the program). As applied to program evaluation, the expected *outcomes* are contingent on (a) the nature of the *program*, (b) who the *participants* are, and (c) what is *measured*. Thus, for each major component of the evaluation—program, participants, measures, and outcomes—there are theoretical and observed patterns, and the degree of correspondence between the theoretical and observed patterns is called the pattern match.

This paper illustrates one type of pattern match—the measurement pattern match—and shows how it can be used as a framework to provide evidence for the construct validity of an evaluation instrument. In order to show evidence for construct validity, it is necessary to have (1) a theory or conceptualization of the expected relationships between the constructs of interest and related constructs from which they must be distinguished (the theoretical measurement pattern); (2) observed interdependence between purported measures of the constructs of interest and of related constructs (the observed measurement pattern); and (3) a "match" between these two patterns (Trochim, 1985).

Requests for reprints should be sent to Jules M. Marquart, Department of Human Service Studies, MVR Hall, Cornell University, Ithaca, NY 14853.

The structured conceptualization method (Trochim, 1986; Trochim & Linton, 1986) was used to develop the pattern of expected relationships among the constructs, as represented by the graphical display or "concept map" from the multidimensional scaling analysis. Measures of the observed relationships were obtained from the data collected on the variables in the program evaluation. The "match" or correspondence between these respective theoretical and observed patterns can then be assessed both visually and quantitatively.

This pattern matching framework for construct validity is applied to an evaluation of a child care program provided by a large medical complex. The study assessed the effects of the employer-sponsored program on employees' work-related attitudes and behaviors, as well as their opinions about other work and family issues.

The framework for this study is provided in Figure 1. The theoretical measurement pattern was developed from the perspective of administrators of hospitals that provide child care assistance to their employees. Administrators were chosen for the theory development because they have a total organizational framework and are usually the decision makers responsible for authorizing such programs. Health administrators, in particular, are a relevant group because hospitals and other health care institutions, more than any other type of company or organization, have been in the forefront of providing child care programs for employees.

The observed measurement pattern was obtained from employees' responses to a questionnaire about the variables of interest. The employees were all working parents with preschool or school-aged children, who used either the employer-sponsored program or other child care arrangements.

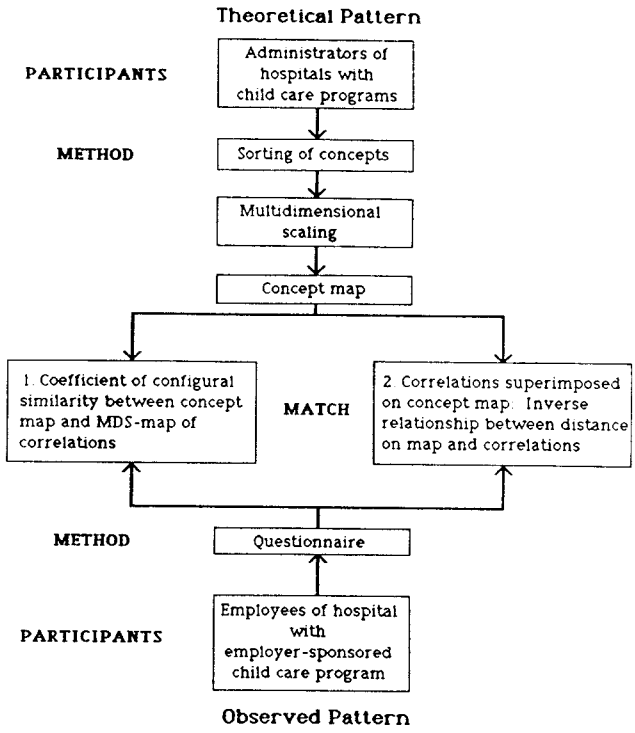


Figure 1. Framework for measurement pattern matches.

The methods for generating the theoretical and observed measurement patterns and for assessing the match between the two patterns are further explained in the next section.

METHODS

Theoretical Measurement Pattern

Nine volunteers were recruited from the health administrators attending the 1986 annual Health Executive Development Program (HEDP).[1] Two of the nine participants in the conceptualization activity were female and the rest were male. They worked for public and private hospitals, and all geographic areas of the country were represented. A wide range of child care assistance was also represented, including on-site centers (3), information and referral services (2), subsidy or voucher programs (1), in-kind contributions—space (1), and a combination of assistance (2). The majority of the participants considered themselves somewhat or very familiar with issues related to employer-sponsored child care.

[1]HEDP is an annual professional conference of health administrators sponsored by the Sloan Program in Health Administration, Department of Human Service Studies, Cornell University.

Concepts. The concepts used in the conceptualization activities were selected by the evaluator based on the variables investigated in the child care program evaluation. They included attitudinal, behavioral, and organizational concepts related to employer-sponsored child care. Specifically, they were: (a) *attitudinal:* recruitment, job satisfaction, organizational commitment, satisfaction with child care arrangement, stress in balancing work and family, and stress regarding child care; (b) *behavioral:* absenteeism and turnover; and (c) *organizational or structural:* organization's personnel policies, supervisor's personnel practices, and employer-sponsored child care.

Data collection and analysis. Two data collection sessions were held with the administrators: the first at the beginning of the conference, and the second, one week later.

At the beginning of the first session, the investigator

explained that she was interested in obtaining their opinions and ideas about employer-sponsored child care programs from their perspectives as administrators in hospitals that sponsor or have considered such programs. A sorting procedure (Rosenberg & Kim, 1975) was used to obtain data on how the health administrators perceived relationships among the concepts, and was adapted from the structured conceptualization method (Trochim, 1986; Trochim & Linton, 1986). Each volunteer was given a list of the concepts along with an envelope containing a set of eleven 3″ × 5″ index cards, with each concept name typed on one card. The investigator read through the list of concepts, and asked if there were any questions about the meaning of any of them. She clarified for one participant that all the attitudinal and behavioral concepts referred to the employee (or were employee-level) while the others were organizational-level.

The participants were then instructed to sort their cards into piles based on which ones seemed to "go together." They were told that this similarity could be based on any factors that seemed important to them; also, the number of piles was up to each of them, as long as they made more than one pile and less than eleven. When finished, each volunteer clipped each pile together, and put all the piles back into the envelope with his/her name.

At this point, the investigator asked if any other important concepts seemed to be missing. Productivity was suggested by one participant, although another group member pointed out that productivity may encompass some concepts already on the list, for example, not being absent, being satisfied with your job, and so on.

The card sorting exercise was repeated at a second session one week later. The same directions were given verbally, and the participants quickly sorted their cards and returned the envelopes. Several participants commented that the sorting task seemed easier and faster the second time because they were more familiar with the concepts and procedure. A microcomputer software program called the Concept System (Trochim, 1986) was used to enter and analyze the sort data. The algorithm for the nonmetric multidimensional scaling is contained in the program to produce the pictorial representation of the sort data, which is called a concept map (see Trochim, this volume).

Observed Measurement Pattern

The evaluation summarized here was of a child care program provided by a large medical complex in the midwest. The employer-sponsored program has two components: an on-site center, and an information and referral service to a network of family day care homes that are screened by the center staff.

A quasi-experimental design was used to compare the two groups of employee users of the company-sponsored child care program and two other groups of employees – one with preschool children and the second with school-age children – who were not using the company program. A survey questionnaire concerning employees' child care arrangements and attitudes about work and family issues was sent to all study participants approximately one year after the employer-provided program began.[2]

A correlation matrix was obtained from the survey data using Pearson Product-Moment correlations. The correlations were from the total sample of respondents ($n = 346$) to the survey questionnaire. The overall sample is comprised of technical and professional employees in their thirties. Roughly three-fourths of the respondents are female, and 85% are married. Just over 70% are full-time hourly or salaried employees. The predominance of female employees and the presence of a relatively high percentage of part-time employees are characteristic of hospitals and make them somewhat different than other employers.

Measurement Pattern Match

Because the concept map represents perceived relationships among the concepts and correlation coefficients are a way to measure the strength of relationships, it seems reasonable to link the correlations from the evaluation data with the concepts on the map to assess the match between the theoretical and operational domains. Two ways of assessing this measurement pattern match are then possible.

First match. It is possible to produce a map of the evaluation data by conducting a multidimensional scaling (MDS) analysis of the correlations between the variables on the questionnaire, and then to assess the pattern match between the concept map from the health administrators and the MDS-map of the correlation matrix. A visual, as well as a quantitative, assessment of the similarity between the two maps can be made. For the quantitative assessment, a correlation coefficient can be calculated from the distances between the concepts on the two maps (see Davis, this volume). This correlation is called here the coefficient of configural similarity, and represents the degree to which distances between points on the two maps are similar. A higher coefficient would indicate greater spatial correspondence and, thus, greater configural similarity between the theoretical and observed measurement patterns.

For the operationalization of this pattern match, a

[2]Data on absenteeism and turnover were collected from personnel records but are not included in these analyses (see Marquart, 1988).

concept map was obtained with the eight attitudinal and organizational variables for which correlations could be calculated from the questionnaire. This concept map was produced by conducting a multidimensional scaling analysis of only those eight variables from the two sortings by the administrators. A correlation matrix was then obtained on those variables from the questionnaire, and a map was produced using the absolute values of the correlations. These values were used, just as the sort data, as input for the Concept System to produce the map. For the qualitative assessment of this pattern match, a visual comparison of the placement of the concepts on the two maps was made. For the quantitative assessment, the distances between points on the two maps were correlated.

Second match. Using the principles of convergence and discrimination from the multitrait-multimethod approach to construct validation (Campbell & Fiske, 1959), one might expect that the variables that are located closer together on the concept map would be more highly correlated, or conversely, the greater the distance between variables, the lower the correlation would be. That is, there should be an inverse relationship between distance on the map and the correlations from the evaluation data.

To test this idea, the correlations between each pair of variables from the questionnaire were superimposed on the concept map. A visual assessment was then made of the expected inverse relationship between distance on the map and the strength of the correlations.

RESULTS

This section presents (a) the perceived relationships among the concepts related to employer-sponsored child care as portrayed in a concept map (theoretical pattern); (b) the correlations among the variables obtained from the survey questionnaire (observed pattern); and (c) the match or degree of correspondence between the theoretical and observed patterns.

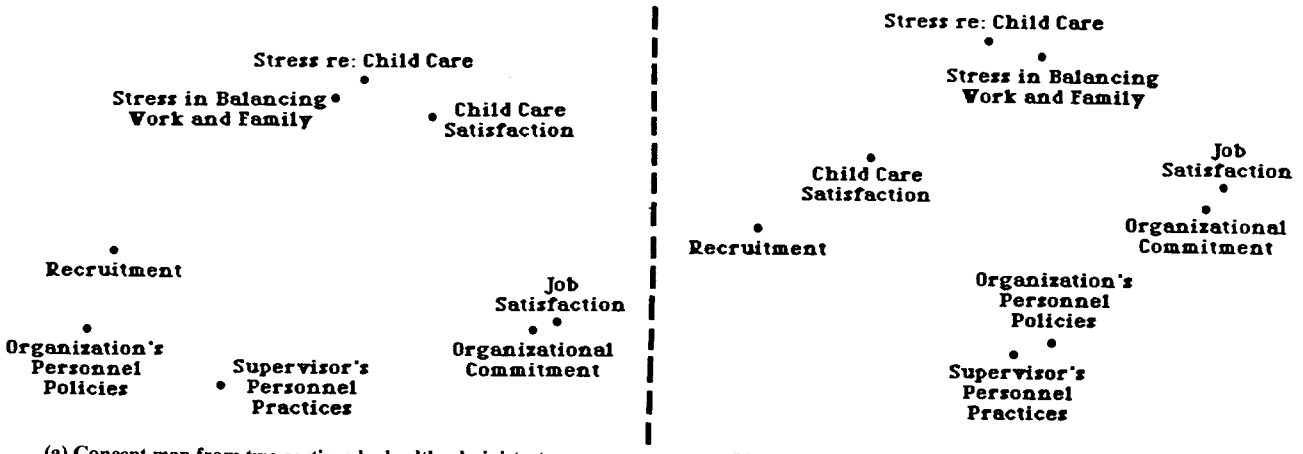
Theoretical Pattern

The concept map shown in Figure 2a depicts the results of the multidimensional scaling analysis of the attitudinal and organizational concepts by the health administrators. This map is a two-dimensional solution, which is chosen for its ease of interpretation, and represents the theoretical measurement pattern. The horizontal dimension is exaggerated by the plotting routine; however, the actual coordinate values are preserved.

Since location on the map is a function of perceived similarity, the concepts that were most frequently sorted together are closest to each other, and those that were seldom or never sorted together are farthest apart or on opposite sides of the map. Based on their similarity, three groups of concepts stand out on the map: one with the concepts referring to work attitudes (job satisfaction and organizational commitment), a second with the concepts pertaining to child care issues (the two stress concepts and child care satisfaction), and a third linking recruitment with personnel policies and practices.

Observed Pattern

The correlation matrix of the variables from the survey questionnaire is presented in Table 1. This matrix represents the observed measurement pattern. The absolute values of the coefficients are used in assessing the pattern match because the strength of the relationship is of primary interest in this case.



(a) Concept map from two sortings by health administrators

(b) Multidimensional scaling map of correlations from questionnaire.

Figure 2. First measurement pattern match: Coefficient of configural similarity between theoretical (a) and observed (b) patterns, $r = .76$.

TABLE 1
OBSERVED MEASUREMENT PATTERN: CORRELATION MATRIX OF VARIABLES FROM QUESTIONNAIRE

Variables	Rec.	Job Satis.	Org. Comm.	Stress re: Child Care	Stress in Bal. Work & Family	Org.'s Pers. Pol.	Supr.'s Pers. Prac.	Satis. with C.C.Arr.
Recruitment	1.00							
Job Satisfaction	.14	1.00						
Organizational Commitment	.10	.52	1.00					
Stress re: Child Care	.15	−.08	−.11	1.00				
Stress in Balancing Work & Family	.13	−.24	−.14	.44	1.00			
Organization's Personnel Policies	.17	.16	.21	−.03	−.05	1.00		
Supervisor's Personnel Practices	.16	.19	.12	.01	−.03	.59	1.00	
Satisfaction with Child Care Arrangement	.11	.08	.10	−.23	−.21	.18	.09	1.00

Pattern Match

First match. A multidimensional scaling analysis was conducted on the correlational data to produce the map shown in Figure 2b. This map represents another way of depicting the observed measurement pattern. The map showing the theoretical pattern (Figure 2a) is placed beside the observed pattern in Figure 2b for visual comparison.

The map from the evaluation data in Figure 2b shows three distinct groups of variables—those pertaining to stress, work attitudes, and personnel issues; the two other variables—recruitment and child care satisfaction—are separated from the others. The pairs of variables that were most highly correlated with each other separated into their own groups. Child care satisfaction, which was most strongly correlated with the two stress variables and organization's personnel policies, was placed at roughly an equal distance from those concepts. Recruitment, which was correlated about equally with the rest of the variables, was placed on the far left of the map.

The coefficient of configural similarity, or spatial correspondence, between the revised theoretical map and the map obtained from the evaluation data was .76. This correlation indicates that there is a relatively high degree of configural similarity between the two maps, that is, between the perceptions of the health administrators of the relationships between the concepts and the actual correlational pattern found in the evaluation data.

Second match. Figure 3 shows the correlations superimposed on the concept map from the administrators. In order to visually differentiate the strength of the correlations, lines of different widths are used—the widest line for the highest values, a medium line for moderate values, and a thin line for the lowest values. This figure represents another way of portraying the correspondence between the theoretical and observed measurement patterns.

In general, the variables that are located closer to each other on the map are more highly correlated, and the variables that are located farther apart have lower correlations. For example, the three most highly correlated pairs of variables are organization's personnel policies and supervisor's personnel practices ($r = .59$), job satisfaction and organizational commitment ($r = .52$), and stress related to child care and stress in balancing work and family ($r = .44$). For the most part, these three pairs of variables are located closer to each other than to any other variables. Also, the three variables within the group pertaining to child care issues are more highly correlated with each other than with variables in the other two groups. Conversely, the two personnel variables have the lowest correlations with the two stress variables, and these two sets of variables are located farthest apart on the map (taking into account the artificial elongation of the map).

The only exceptions to this confirmatory pattern of relationships are between job satisfaction and stress in balancing work and family ($r = .24$), and between organizational commitment and personnel policies ($r = .21$). The placement of recruitment on the map is also interesting to note. Although recruitment is correlated rather weakly and about equally with all the other variables, it is most highly correlated with organization's personnel policies (.17) and supervisor's personnel practices (.16), and recruitment is also located closer to those two variables on the map than to any others.

The inverse relationship between distance on the map and the correlations can be observed in this figure. In general, the greater the distance between variables, the lower the correlation was, and vice versa. Overall, a fairly high degree of correspondence was found between the administrators' perceptions of the relationships between the concepts related to employer-sponsored child care and the correlations among the variables on the questionnaire.

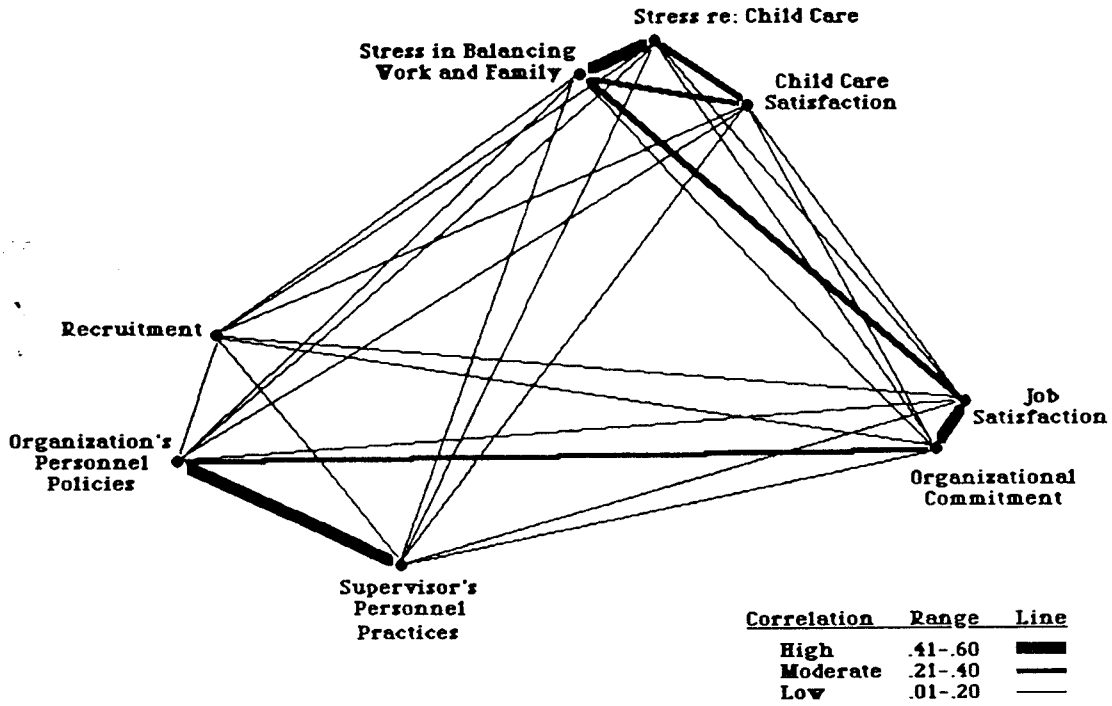


Figure 3. Second measurement pattern match: Evaluation correlations superimposed on concept map.

CONCLUSIONS

A fairly high degree of correspondence was found between the health administrators' perceptions of the relationships between the concepts related to employer-sponsored child care and the correlations from the child care program evaluation, that is, between the theoretical and observed measurement patterns. This claim is supported both by the high coefficient of configural similarity ($r = .76$), and by the inverse relationship found between the distances between the concepts on the map and the correlations. These two successful matches between the theoretical and observed measurement patterns provide substantial evidence for construct validity of the survey instrument.

Although this study represents a first step in using the pattern matching approach to assess the correspondence between theory and data, it suggests other ways in which such a framework might be used in program evaluation. The approach could be used to guide the entire instrument development and validation process by involving the program constituents in conceptualizing the major concepts of interest, using those concepts to construct an instrument, and then using the data collected to provide evidence for the construct validity of the instrument as well as to assess program outcomes (for example, see Marquart, 1987, and Valentine, this volume.)

On the methodological side, the approach offers an alternative to methods commonly used to assess the validity of an instrument. It has an advantage over factor analysis for scale development in that multidimensional scaling does not impose the same constraints on the level of measurement. As was illustrated in the examples, because of the flexibility of the multidimensional scaling procedures, the sorting data as well as the scaled data from the questionnaire could be analyzed and compared. With further application, the pattern matching approach will be refined so that its potential for improving the development and use of theory in program evaluation can be determined.

REFERENCES

CAMPBELL, D.T. (1966). Pattern matching as an essential in distal knowing. In K.R. Hammond (Ed.), *The Psychology of Egon Brunswick*. New York: Holt, Rinehart and Winston.

CAMPBELL, D.T., & FISKE, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

MARQUART, J.M. (1987, October). *Concept mapping: A tool for defining the conceptual domain.* Paper presented at the Annual Meeting of the American Evaluation Association, Boston.

MARQUART, J.M. (1988). *A pattern matching approach to link program theory and evaluation data: The case of employer-sponsored child care.* Doctoral dissertation, Cornell University, Ithaca, NY.

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

TROCHIM, W.M.K. (1985). Pattern matching, validity, and conceptualization in program evaluation. *Evaluation Review, 9,* 575-604.

TROCHIM, W.M.K. (1986). *Structured conceptualization: A resource guide and program report.* Unpublished manuscript, Cornell University, Department of Human Service Studies, Ithaca, NY.

TROCHIM, W.M.K. (in press). Pattern matching and program theory. *Evaluation and Program Planning.*

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

Evaluation and Program Planning, Vol. 12, pp. 45-52, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

STRUCTURED CONCEPTUALIZATION

A Framework for Interpreting Evaluation Results

VALERIE J. CARACELLI

Cornell University

ABSTRACT

The structured conceptualization method can be used as a descriptive and interpretive tool for understanding results from evaluations and applied social research programs. A case study of women in Fordham University's EXCEL reentry program for adults demonstrates the use of this method in the descriptive and exploratory phase of research. Two groups of women were identified and contrasted on the California Q-set which contains 100 personality descriptors. ANCOVAs revealed few statistically significant differences between Integrators, women with focused career goals, and Seekers, women with unfocused career goals. However, when differences between Integrators and Seekers on each of 100 personality items are overlaid on a map of these items which was generated by the combined use of multidimensional scaling and cluster analysis procedures, a pattern is observed which is readily interpretable and indicates that these two groups are not homogeneous. Similarly, women who completed their first year of college were contrasted on all 100 personality items with women who did not complete their first year of study. The method of structured conceptualization is viewed as useful in helping evaluators make sensible inferences from findings even when statistically significant differences are not evident because of low statistical power due to small sample sizes or inadequate measures. Further, the method can serve as a tool for examining and strengthening statistical conclusion validity when multiple significance tests are used.

Structured conceptualization is an innovative research method developed in part to enhance "pattern matching" of conceptual and operational domains in the field of program evaluation (Trochim, in press). The emphasis in this paper will be on the use of the structured conceptualization method as a descriptive and interpretive tool for understanding the results from evaluation studies and applied social research programs. Here, concept maps are used in a pattern matching perspective by overlaying estimates of differences between defined groups onto a concept map of personality characteristics. The maps generated by the structured conceptualization procedure offer a graphic representation of the obtained pattern of results and serve as a framework for the display of statistical data. This paper discusses the potential use of such maps for helping us discern meaningful inferences from our data even when univariate statistical tests do not reach statistical significance.

In order to demonstrate the utility of the conceptualization method an analysis of a portion of data obtained from the first wave of a mixed-methods longitudinal study will be presented. The study was set within an ecological framework (Bronfenbrenner, 1979) and examined the experience of women reentering the educational system during the initial period of adjustment from the time just prior to college entrance to the end of their first academic year. By their nature role transitions pose potential challenges since they disrupt established patterns of behavior and require personal adjustment (George, 1980; Jacobi, 1987). Therefore, an

I would like to thank Bill Trochim for sharing his expertise as mentor and collaborator in the analysis presented. My gratitude is also extended to Stephen Hamilton and Daryl Bem who have been mainstays in all aspects of this research project.

Requests for reprints should be sent to Valerie Caracelli, Department of Human Development and Family Studies, MVR Hall, Cornell University, Ithaca, NY 14853.

inquiry was made to determine what psychological changes, if any, occurred in the lives of women who were making this transition as adults. Another purpose of this study was to provide teachers, counselors, and administrators in higher education with a description of the personal characteristics of women who differed along motivational dimensions and expectations in beginning or returning to college. It was anticipated that women who differed in orientation might respond differently to the college experience.

METHOD

Sample

All participants were enrolled in Fordham University's EXCEL program. The EXCEL program is a "special entry program for adults, leading to a B.A. degree." The establishment of this program in 1970, with its interdisciplinary focus and credit for life experience, represented a major effort on the part of the University to adapt their traditional liberal arts programs to the needs of adult students by providing a bridge-program to college.

The sample of 70 women is fairly representative of entering classes: mean age 34, 30% minority women, majority attending evening classes. Most of these women were in the labor force (77%) or intended to resume employment. Forty percent of the women lived with a spouse or mate and of these about one-third had children at home. Single mothers with children at home were approximately 16% of the sample. The remaining women were independent (without mate or children).

Design

A pre and posttest academic year longitudinal design was used to document the early transition experience of adult women entering or returning to college. Interviews were conducted with participants prior to college entrance and at the end of their first year. A number of paper and pencil measures were also given. This paper will focus on one of the instruments, the California Q-set, which was given to participants at both points of data collection in order to obtain a broad range of personality characteristics and to serve as a means of describing subgroups within the sample. (See Caracelli, 1988 for a more complete description of the methods and results of this study.)

Measures

All participants completed the California Q-set which is a standardized personality measure that has been used extensively in psychological research (Bem & Funder, 1978; Block, 1961). The Q-set contains 100 descriptive personality statements (e.g. critical, skeptical, has insight into own motives and behavior, etc.). In this procedure the universe of items is essentially predefined for participants. Using a forced symmetric distribution each woman sorts the cards into 9 categories ranging from 1 (those items that are least characteristic) to 9 (those items that are most characteristic of her personality). Thus each item receives a score from 1 to 9; the distribution has a mean of 5 and a standard deviation of 2.

Development of the Conceptual Map

There were three steps involved in the development of the conceptual map. First, an individual multidimensional scaling (MDS) analysis (Davison, 1983; Kruskal & Wish, 1978) was conducted using the INDSCAL model. A two-dimensional solution was selected primarily for the ease of interpretation and comprehension it affords. Kruskal & Wish concur that a two-dimensional configuration is more useful than a configuration of 3 or more dimensions when MDS serves as a foundation to display clustering results as is done here. Essentially the MDS procedure scales the 100 items in relation to one another based on the item rating similarities. The output is a two-dimensional snapshot of the conceptual arrangement of all 100 items. More specifically, it is a spatial, map-like representation consisting of a geometric configuration of all 100 points. Second, Ward's hierarchical cluster analysis procedure was used to group the items into conceptual domains of personality. The decision to depict a four cluster solution was again based on ease of interpretation. Cluster names were finalized through a process of consensual agreement among several social science researchers. Table 1 lists the Q-items contained within each of the following four clusters: Affiliative-Intellectual; Extraversion-Sociability; Adaptive-Maladaptive; Self-Assurance. Third, subgroups were identified and differences between these groups on each of the 100 items were examined using an analysis of covariance (ANCOVA) model. These analyses tested for differences between groups on each item at the end of their first year of college, controlling for pretest differences for that item. Bar graphs representing the t-values obtained in these analyses were superimposed on the map for each of the 100 items. Those bars containing five units were significant at the .05 level.

The first subgroup analysis made a distinction between women in the sample on the basis of their career goals. Two types of reentry women were identified through a qualitative analysis of the interview transcripts (see Patton, 1980 regarding analyst-constructed typologies). One group, women with focused career goals were classified as Integrators. Generally these women had specific professional goals (e.g., becoming

TABLE 1
Q-ITEMS BY CLUSTER

Self Assurance
- Q1. Critical, skeptical
- Q16. Introspective
- Q31. Regards self as attractive
- Q39. Unusual, unconventional thinking
- Q52. Assertive
- Q57. Interesting, arresting
- Q62. Rebellious, non-conforming
- Q69. Sensitive to demands
- Q71. High aspiration level
- Q74. Self-satisfied
- Q77. Appears straightforward, candid
- Q81. Physically attractive
- Q82. Fluctuating moods
- Q83. Sees to heart of important problems
- Q90. Concerned with philosophical problems
- Q91. Power oriented; values power
- Q92. Social poise
- Q98. Verbally fluent

Affiliative-Intellectual
- Q2. Dependable, responsible
- Q3. Wide range of interests
- Q4. Talkative
- Q5. Giving toward others
- Q8. Intelligent
- Q11. Protective of those close to her
- Q15. Skilled in imaginative play, humor
- Q17. Sympathetic, considerate
- Q18. Initiates humor
- Q20. Rapid personal tempo
- Q24. Prides self on being objective, rational
- Q26. Productive
- Q28. Arouses liking, acceptance
- Q29. Turned to for advice, reassurance
- Q35. Warm, capacity for close relationships
- Q43. Facially/gesturally expressive
- Q51. Values intellectual & cognitive matters
- Q56. Responds to humor
- Q58. Enjoys sensuous experiences
- Q60. Has insight into own motives & behavior
- Q66. Aesthetically reactive
- Q70. Ethically consistent
- Q84. Cheerful

Extraversion-Sociability
- Q7. Conservative values
- Q10. Anxiety finds outlet in bodily symptoms
- Q19. Seeks reassurance
- Q21. Arouses nurturance
- Q32. Aware of impression she makes
- Q33. Calm, relaxed
- Q44. Evaluates motivation of others
- Q46. Daydreams, fantasizes
- Q54. Gregarious
- Q59. Concerned with bodily functioning
- Q64. Socially perceptive
- Q72. Concerned with own adequacy
- Q75. Consistent, clear-cut personality
- Q79. Ruminates, pre-occupying thoughts
- Q80. Interested in opposite sex
- Q85. Emphasizes action, non-verbal behavior
- Q88. Personally charming
- Q89. Compares self to others
- Q93. Feminine
- Q95. Offers advice
- Q96. Values independence, autonomy

Adaptive-Maladaptive
- Q6. Fastidious
- Q9. Uncomfortable with uncertainty, complexity
- Q12. Self-defensive
- Q13. Thin skinned
- Q14. Submissive
- Q22. Feels lack of personal meaning
- Q23. Extrapunitive, projects blame
- Q25. Over-controlled
- Q27. Condescending
- Q30. Gives up in the face of frustration
- Q34. Over-reactive to frustration, irritable
- Q36. Negativistic; obstructs, sabotages
- Q37. Guileful, manipulative, opportunistic
- Q38. Has hostility
- Q40. Vulnerable to threat, fearful
- Q41. Moralistic
- Q42. Delays or avoids action
- Q45. Brittle ego-defenses
- Q47. Feels guilty readily
- Q48. Keeps people at a distance
- Q49. Basically distrustful of others
- Q50. Unpredictable, changeable
- Q53. Unable to delay gratification, uncontrolled
- Q55. Self-defeating
- Q61. Creates and exploits dependency
- Q63. Judges self/others in conventional terms
- Q65. Stretches limits
- Q67. Self-indulgent
- Q68. Basically anxious
- Q73. Eroticizes situations
- Q76. Projects feelings/motivations onto others
- Q78. Feels cheated/victimized; self-pitying
- Q86. Repressive
- Q87. Complicates situations
- Q94. Expresses hostility directly
- Q97. Emotionally bland; flat affect
- Q99. Self-dramatizing; histrionic
- Q100. Does not vary roles

a lawyer or child psychologist) and viewed college as providing a means to those ends. The other group, women with unfocused career goals, were classified as Seekers. These women were concerned about career issues but were less directed and not committed to specific career goals. For the most part, they expressed feelings of being "in a waiting period," "marking time," "trying to leave things open." They viewed college as a tool of exploration.

The second subgroup analysis distinguished women

in the sample in terms of their persistence in pursuing their college education. The relatively few ($n = 7$) women who withdrew from the EXCEL program before the end of the first year were classified as Nonpersisters while those who completed their first year were identified as Persisters.

RESULTS

One purpose of this research was to provide a description of these reentry women on the basis of personality characteristics. Figure 1 shows the concept map for the 100 Q-set items. In general, items which are closer together on the map had similar ratings while those which are farther apart had more discrepant ones. We can begin a rudimentary interpretation of the personality items using Figure 2 which combines the results of the MDS and cluster analysis procedures. This figure is essentially a graphic portrayal of the cluster results reported in Table 1. When we examine the results for each item, we find that the items rated as "most" or "quite" characteristic by women in the sample ($n = 50$) tend to be located on the left side of the map. In fact 85% of these items are in the Affiliative-Intellectual cluster. This finding is somewhat surprising given that the sex-differences literature often separates characteristics in this cluster into two separate and opposing constructs. It is clear that these reentry women feel they have a mix of traits associated with these theoretically separate constructs (see Emmerich, 1973 and Gilligan, 1982 for thoughtful appraisals of the research surrounding this issue). Overall, women in the sample are describing themselves as dependable, responsible, and productive. On a cognitive level they perceive themselves as intelligent, valuing intellectual and cognitive matters, having a wide range of interests, high aspiration level, and as aesthetically reactive. On an interpersonal level they feel they respond to humor, are warm, sensuous, and considerate of others.

On the right side of the map are the Q-items that these women rated as "least characteristic" or "quite uncharacteristic" of themselves. Almost all of these items are in the Adaptive-Maladaptive cluster. In general, women in the sample do not perceive themselves as deceitful, manipulative, or as creating and exploiting dependency in others. In response to life's vicissitudes they do not feel they are emotionally bland or self-pitying. Nor do they see themselves as giving up or withdrawing in the face of frustration or adversity. It appears the women strike a balance in valuing interpersonal and instrumental aspects of their lives and reveal their own perceptions of inner strength in meeting challenges such as the one posed by their recent transition to college.

Subgroup Analysis: Integrators versus Seekers

It was anticipated that some psychological changes might occur for women making a transition of this nature as adults, and further that women who held different motivations and expectations when returning to college might respond differently to the college experi-

Figure 1. Two-dimensional concept map showing the 100 Q-set items.

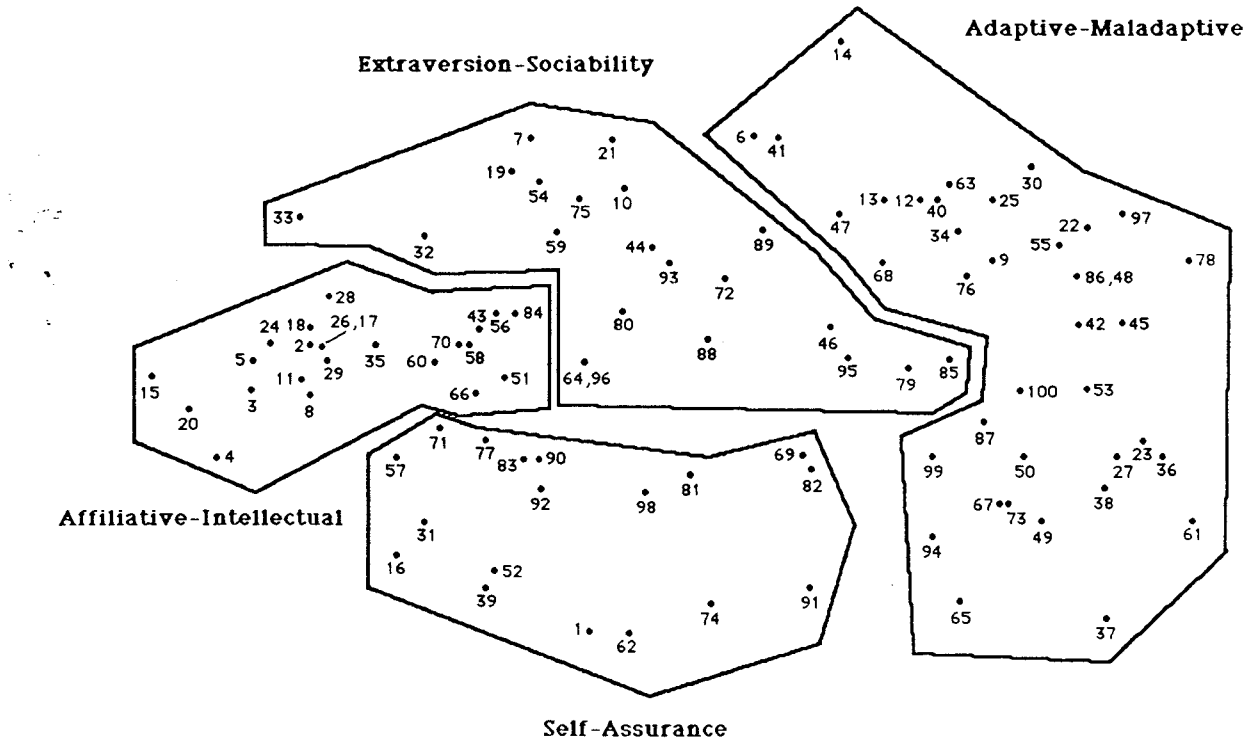


Figure 2. Two-dimensional concept map showing individual Q-set items, clusters, and cluster names.

ence. Figure 3 depicts the results of the ANCOVAs that compare the personality characteristics of these women at the end of their first year of college.

Only women who retained the same classification for pre and posttest are included in this analysis (Integrators $n = 19$, Seekers $n = 22$). Higher bars on the graph indicate higher estimates of group differences (i.e., t-values) for that Q-item while the shading of the bars indicates which of the two groups scored higher in mean value. Often we are constrained in reporting results of studies to emphasize only the findings that are statistically significant. That being the case there would be little to say about these two groups of women since significant differences were few in number.

What do we gain by having a conceptual map as a framework for reporting our findings? In Figure 3 we can readily detect a pattern across the four clusters. The clusters on the left side of the map show a preponderance of dark columns. The ANCOVAs reveal that by the end of their first year in college Integrators tended to score themselves higher than Seekers on these Q-items. Thus, the map clearly shows that Integrators rated themselves higher on almost all of the cognitive characteristics in the Affiliative-Intellectual cluster. These are characteristics we would expect to be important in a college environment, e.g., perceiving oneself as intelligent, valuing intellectual and cognitive matters, having a wide range of interests, and being skilled in imaginative play.

On the right side of the map we see that the reverse is true. The clusters show a preponderance of white columns indicating that Seekers tended to rate themselves higher on these Q-items. Again, the map clearly shows the Q-items that distinguish between the groups. In the Adaptive-Maladaptive cluster the t-values show that it is more characteristic of Seekers to perceive themselves as basically anxious and self-defeating. In comparison to Integrators they see themselves as more likely to delay action, be unpredictable, and to be concerned with their own adequacy.

A typical interpretation of these data would lead one to conclude that Integrators and Seekers do not differ since statistical differences between the groups using the ANCOVA model were not greater than we might expect by chance. In this example, even a multivariate ANCOVA would be likely to yield a nonsignificant overall F-value. Nevertheless, by overlaying these results on a mapping background it is reasonable to conclude that the overall pattern suggests an interpretable finding regarding how Integrators and Seekers may differ. The conceptual map effectively enables us to draw portraits of the two groups of women. We can place greater confidence in the characteristics that statistically distinguished Integrators and Seekers (5 unit t-value bars) because the obtained pattern of differences supports these findings. Further, the pattern suggests we might erroneously conclude a lack of true difference between groups when actually low statistical power

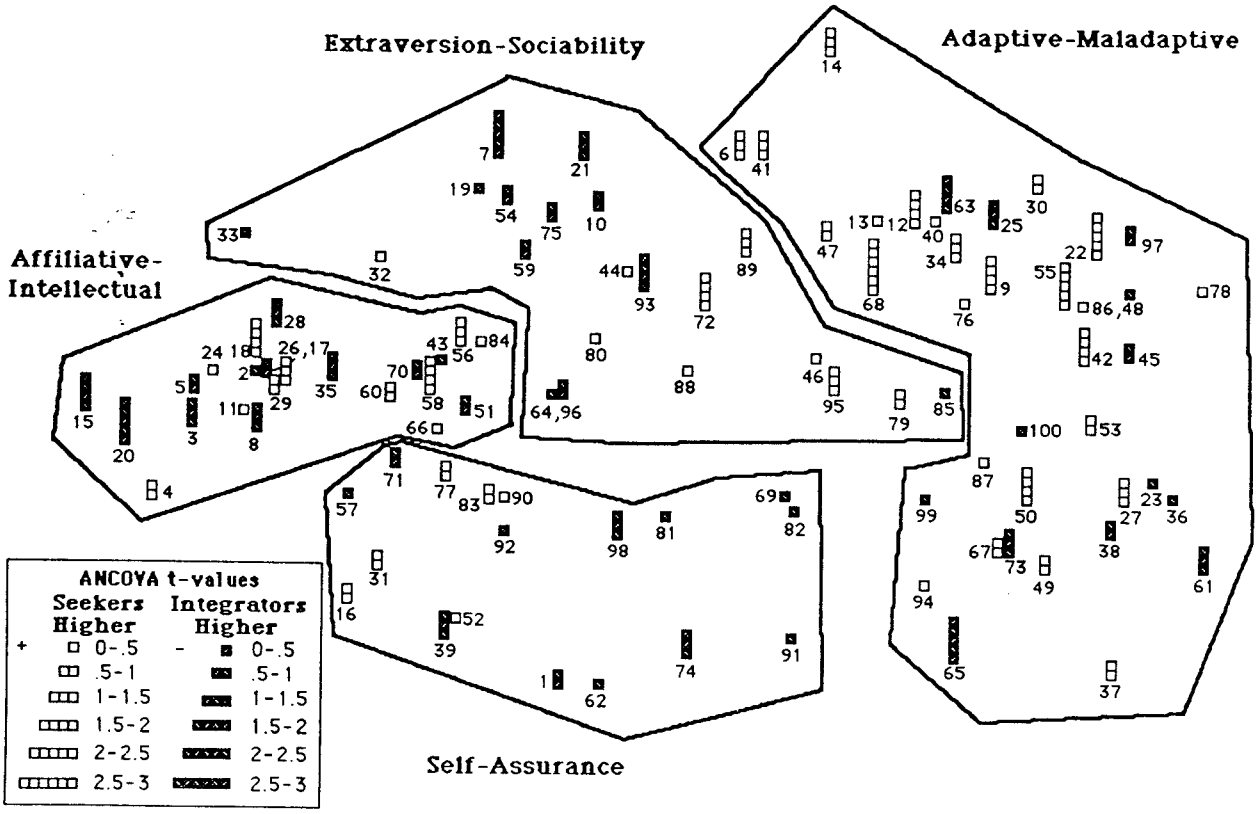


Figure 3. Concept map showing ANCOVA *t*-value results testing for differences between Integrators and Seekers.

may have influenced the outcome of the analyses of covariance.

Subgroup Analysis: Persisters versus Nonpersisters

An even clearer illustration of this line of argument can be demonstrated with the concept map in Figure 4 where the distinction between groups is even more striking. Higher *t*-values are clearly predominant for Persisters in the Self-Assurance cluster, especially for item 71 which relates to high aspiration level. Persisters also rate themselves more highly on the cognitive traits contained in the Affiliative-Intellectual cluster. In the Adaptive-Maladaptive cluster the *t*-values indicate that in contrast to Persisters it is more characteristic of the Nonpersisters to perceive themselves as submissive, self-indulgent, unable to delay gratification, and giving up and withdrawing in the face of frustration. Again, it is important to note that the statistical significance of individual Q-items, where it occurs, is greatly supported by the pattern of differences between groups that is apparent in Figure 4.

Persistence is a complex phenomenon of interest to counselors and educational administrators. These results must be viewed in the context of the overall study. The ostensible reasons given by these women for withdrawing from the program revealed concerns about finances, job transitions, and family issues. Nonpersisters enter college with lower aptitude scores and the personality profiles we derive from the conceptual map indicate that compared to persisters these women reveal less academic motivation and a general mode of responding to adverse conditions that may warrant targeted intervention strategies on the part of college counseling services. In this instance, in conjunction with other findings, the conceptual map helps to broaden our understanding of the multifaceted issue of persistence.

CONCLUSIONS

At the outset of this paper the point was made that the structured conceptualization method could serve as a descriptive and interpretive tool for understanding results from evaluation studies, and further, that the conceptual maps generated by the multidimensional scaling and cluster analysis procedures could serve as a framework for the display of statistical data. There are several concluding points that need elaboration concerning the utility of this approach.

First, the first phase of the study that was presented

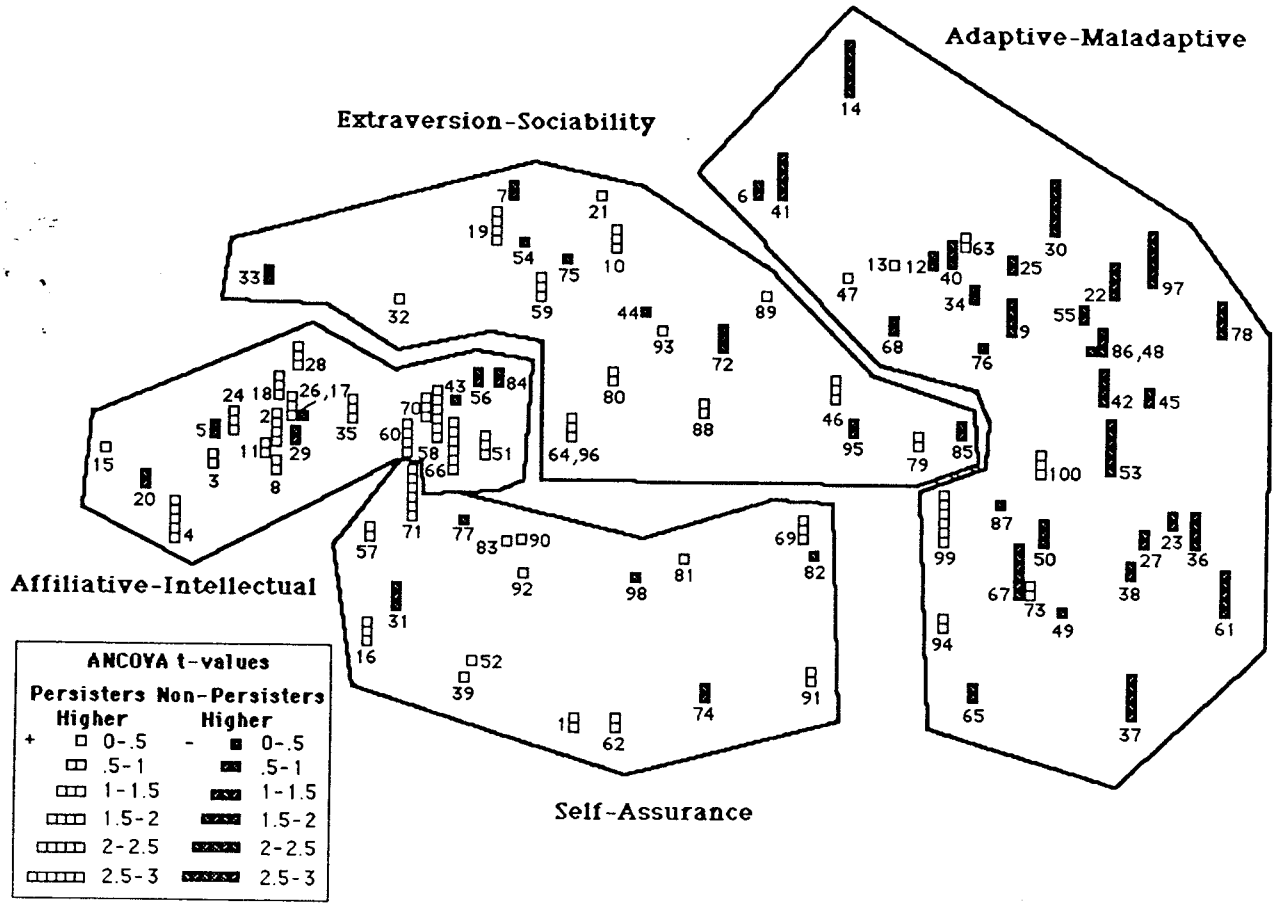


Figure 4. Concept map showing ANCOVA *t*-value results testing for differences between Persisters and Non-Persisters.

as an illustration was intended to be descriptive and exploratory and the conceptual maps were interpreted here with that purpose in mind. In studies designed on the basis of pre-existing theory, the structured conceptualization method can also serve as a tool of confirmation by using pattern matching techniques that contrast obtained patterns derived from empirical data with expected patterns based on theory (Trochim, in press; Trochim, 1985). The similarity or match between patterns determines the degree to which we can claim construct validity for the cause-effect relationship. In this study, for instance, it would have been interesting to generate a theoretical pattern (i.e., concept map) of the Q-items based on a panel's judgments of these items' conceptual similarity. Estimates of group differences could then be overlayed onto this theoretical measurement pattern.

Second, in evaluations where the sample size is small or where measures may not be adequate we may fail to find statistical differences simply due to low power. In this case we can look at the pattern of non-significant findings and make some determination of where we might expect to find differences if we had more power in our study to pick up those effects. We also need to keep in mind that differences do not have to be statistically significant to be meaningful and that even statistically significant results may lack practical significance.

Third, in evaluations requiring multiple significance tests we face the "fishing expedition" or "error rate" problem—that we are likely to suspect that some proportion of the individual tests are significant by chance alone. Mark and Cook (1984) outline several corrective statistical adjustments that can be made. They note, however, that despite precautions it is still possible to conclude that a treatment is related to an outcome when it is not (Type I error) or that a treatment and outcome are not related when in fact they are (Type II error). The structured conceptualization method can serve to examine the error rate problem graphically since we would expect that if obtained differences were due to "chance" they would be spread throughout the conceptual map in some random fashion. The map can serve as a line of defense against the "chance" argument since by visual inspection we can judge whether the differences between groups are random with respect to the interrelational structures of the

measures. We can do this by placing these significant differences in the context of the surrounding pattern of results.

In essence, the argument for the utility of the structured conceptualization method outlined here is analogous to the discovery of the Nazca lines in Peru. At ground level these lines on the desert plain of Nazca appear to be nothing more than furrows in the ground. When viewed from the air the lines form definite outlines of birds, animals, human and geometric figures traced by Pre-Incaic civilization. The maps illustrated in this paper provide such an aerial view. The display of statistical data in this fashion has the potential of elucidating patterns in our data that allow us to reduce the level of inferential equivocality and improve interpretation of our findings.

REFERENCES

BEM, D.J., & FUNDER, D.C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review, 85*, 485-501.

BLOCK, J. (1961). *The Q-sort method in personality assessment and psychiatric research.* Springfield, IL: Charles C. Thomas.

BRONFENBRENNER, U. (1979). *The ecology of human development: Experiments by nature and design.* Cambridge, MA: Harvard University Press.

CARACELLI, V.J. (1988). *Adult women in college: Stability and change in identity and personality.* Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

DAVISON, M.L. (1983). *Multidimensional scaling.* New York: Wiley.

EMMERICH, W. (1973). Socialization and sex-role development. In P.B. Baltes & K.W. Schaie (Eds.), *Life-span developmental psychology: Personality & socialization.* New York: Academic Press.

GEORGE, L.K. (1980). *Role transitions in later life.* Monterey, CA: Brooks/Cole.

GILLIGAN, C. (1982). *In a different voice: Psychological theory and women's development.* Cambridge, MA: Harvard University Press.

JACOBI, M. (1987). *Stress among re-entry women students: A contextual approach.* Paper presented at the Annual Conference of the Association for the Study of Higher Education, February.

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

MARK, M.M., & COOK, T.D. (1984). Design of randomized experiments and quasi-experiments. In L. Rutman (Ed.), *Evaluation research methods: A basic guide* (2nd ed.). Beverly Hills, CA: Sage.

PATTON, M.Q. (1980). *Qualitative evaluation methods.* Beverly Hills, CA: Sage.

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

TROCHIM, W.M.K. (in press). Pattern matching and program theory. *Evaluation and Program Planning.*

Evaluation and Program Planning, Vol. 12, pp. 53-57, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

CONCEPT MAPPING FOR PLANNING AND EVALUATION OF A BIG BROTHER/BIG SISTER PROGRAM

PATRICK F. GALVIN

Cornell University

ABSTRACT

This paper describes how concept mapping was used to organize a conceptual framework from which a questionnaire was constructed, in order to evaluate the One-to-One Big Brother/Big Sister youth program. The evaluation results derived from this questionnaire support the claim that the program effectively meets its goals: program participants scored significantly higher on the questionnaire than a comparable group of youth who had not participated in the program. Furthermore, the pattern of results was consistent with the staff's expectations of the program's outcomes as indicated through the concept mapping process.

If the outcomes for social service/educational programs were specific and easily measurable, then evaluating program effectiveness would not be difficult. Most programs, however, provide a variety of services from which participants benefit differently, complicating the evaluation process. Concept mapping assists evaluators in conceptualizing program outcomes, allowing them to design a data collection strategy that strengthens the validity of the results.

This paper describes how the concept mapping technique was used as an integral part of the evaluation process of the One-to-One Big Brother/Big Sister youth program. Concept mapping helped organize a conceptual framework from which a questionnaire was constructed. The evaluation results derived from this questionnaire support the claim that the program effectively meets its goals: program participants scored significantly higher on a questionnaire than a comparable group of youth who had not participated in the program. Additionally, the pattern of results was consistent with the staff's expectations of the program's outcomes, which further suggests that the program is meeting its goals. In the following sections, the concept mapping process and questionnaire development are described, after which the evaluation results are briefly reviewed.

METHOD

Program

The One-to-One Program Proposal states the purpose of the program in the following terms:

> The One-to-One Family Support Program seeks to address a general need on the part of children for focused, caring, individual adult attention . . . Since changes in society (e.g., more single-parent families, more two-job families, increased loads on teachers and others who work with children) make it less possible for these people to spend focused individual time with children, the need must be met through agency-sponsored programs such as One-to-One . . . Youth need skills, opportunities and supports to enable them to deal with the issues of adolescence . . . Specifically, youth need affective education through schools and community programs to learn and practice the skills needed for effective communication, conflict resolution, decision-making, problem solving, satisfying interpersonal, group and inter-group relationships. (Project Proposal; p 1)

The One-to-One Program staff assert that through structured opportunities, children can develop the physical, social and life skills necessary to function successfully in the community. Matching young people with properly trained and supervised adult volunteers provides youth with these structured opportunities.

Requests for reprints should be sent to Patrick Galvin, Department of Education, Roberts Hall, Cornell University, Ithaca, NY 14853.

State officials requested that the One-to-One Program be evaluated as a condition for continued funding; they wanted evidence of the program's effectiveness. While the program's outcomes are described in terms of the child's social and personal adjustments, the staff recognized that the volunteers' involvement with their Little Siblings consisted mostly of activities such as eating ice cream, hiking, playing baseball and basketball, or doing craft activities, to name a few. Thus it was obvious that a child might benefit in many different ways from participating in the program. Further, within any one of these program activities benefits might manifest themselves differently; one child might identify the program's benefits with an expansion of his or her opportunities, another child might identify the benefits with changes in personal feelings. A third might simply develop increased enthusiasm for a particular activity.

Concept Mapping Program

During preliminary discussions with the nine One-to-One staff members about the nature of the program, long lists of program activities and outcomes were easily generated, creating debate among staff members over which items and outcomes fairly represented the program. Concept mapping provided a method by which both the staff and evaluator could conceptualize the program, enabling them to begin designing an appropriate questionnaire.

The concept mapping process followed the steps outlined in Trochim (this volume) and involved group brainstorming, card sorting, and the interpretation of the resulting map (no ratings of statements were done). The One-to-One staff were able to brainstorm over one hundred items describing the program. Some of these items were very concrete, such as "eating ice cream" or "going to a movie"; others were more abstract, such as "feeling better about one's self" or "an adult friend to talk to." During the session, the staff were encouraged to think of program areas not yet included on the list. By the end of the hour the flow of ideas for this session was exhausted. The staff by and large agreed that the list fairly represented the various aspects of the program.

The second step of the Concept mapping technique—card sorting—helped reveal more about the nature of the program as perceived by the staff. Each brainstormed item was typed out on 3 × 5 cards, and a copy of the entire set of 98 unduplicated items was given to each staff member. Each staff member then sorted all the items into piles of related items, giving each pile a title. The staff found this task enjoyable and easy to do, and, like the brainstorming session, it took less than an hour of their time.

To give the reader a sense of the program, Table 1 lists the titles of sorted items given by three staff members. These sorted piles of brainstormed items formed

TABLE 1
TITLES OF SORTED GROUPS FOR THREE STAFF MEMBERS

Staff Member 1	Staff Member 2	Staff Member 3
Personal growth	Feel good about self	Kids get experience
General fun stuff	Self knowledge	Benefits for parents
Arts & crafts	Intellectual play	Activities
Physical stuff	Physical play	
School academics	Helping with life skills	
Family help	Networking/social service	
	Community service	

the data used to construct the concept map (see Trochim, this volume, for further details about how the concept maps are formed). After consultation with the One-to-One staff, 23 clusters were identified on the concept map as shown in Figure 1, each bound by a common theme: some of them well-defined by the Concept Mapping procedure, others by the rigor of common sense.

While the Concept Map pictorially represents the program, it is not inherently meaningful. Interpretation of the map is organized around the two continuums: an Individual-Group continuum, and an Activities-Social Skills continuum. The Individual-Group continuum differentiates activities in terms of their social organization. The Activities-Social Skills continuum differentiates among characteristics of competence or self esteem. In other words, the basic thrust of the One-to-One Program is to help youth cope personally and socially with dilemmas of adolescence. This is achieved as the youth gain experience and skill participating in a variety of group as well as individual activities. The resulting quadrants represent four conceptual domains of the program: (a) Group/Activity Skills (b) Individual/Activity Skills; (c) Group Social Skills; (d) Individual Social Skills.

One could say that the concept cap represents a local theory about how the One-to-One Program operates. Indeed, when presenting this map to the One-to-One staff, the staff accepted its organization and basic propositions as representative of the program. On the other hand, the reader should note that the map represents an average, if you will, of all the staff members' interpretations of the program through the card sorting technique. Consequently, it will not be surprising to learn that the map did not *fully* satisfy anyone.

Questionnaire Development

The purpose in developing the concept map was to design a questionnaire with outcome measures directly related to One-to-One Program activities. Twenty-three specific dimensions of the program were identified using the concept map. Each question on the questionnaire represented one of the concept map clusters. This

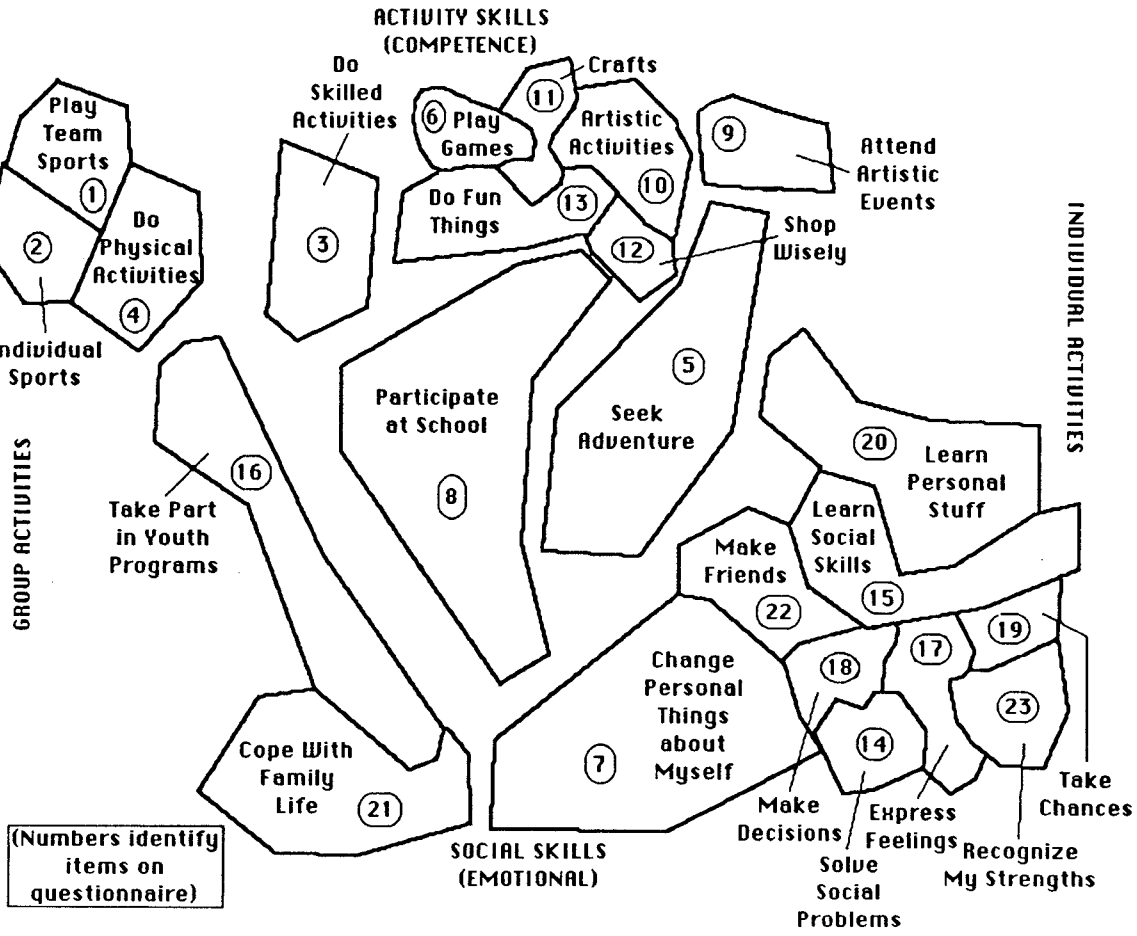


Figure 1. Concept map for the One-to-One Program.

TABLE 2
SAMPLE QUESTIONNAIRE ITEMS

	I GET A CHANCE TO DO THESE THINGS	I AM GOOD AT THESE THINGS	I WANT TO DO MORE OF THESE THINGS
1) PLAY TEAM SPORTS (baseball, soccer, volleyball,	1...2...3...4...5 Not at all Some Yes a lot	1...2...3...4...5 Very bad OK Very Good	1...2...3...4...5 Not at all Some Yes a lot
8) PARTICIPATE AT SCHOOL (finish homework, school	1...2...3...4...5 Not at all Some Yes a lot	1...2...3...4...5 Very bad OK Very Good	1...2...3...4...5 Not at all Some Yes a lot
23) RECOGNIZE MY STRENGTHS (learn what I am good at, special skills,	1...2...3...4...5 Not at all Some Yes a lot	1...2...3...4...5 Very bad OK Very Good	1...2...3...4...5 Not at all Some Yes a lot

ensured that the questionnaire included outcomes, ideas and activities from all relevant program domains.

Additionally, as noted above, participating youth might benefit from the program in a number of different ways. These were categorized as changes in opportunities, feelings about ones self, and attitudes. Hence, a multidimension-multitrait questionnaire was constructed. Table 2 shows three questions used on the questionnaire, and illustrates the format.

The questionnaire was designed for youth ages ten or

older; pretesting established that most youth were able to respond to the questionnaire with little or no guidance.

Design

The One-to-One staff claim that matching youth with a program Big Sibling positively affects their opportunities, feeling and attitudes in a variety of program areas. A quasi-experimental group design was utilized to assess outcomes. The program matched over 140 of the community's youth (including different races, ages, gender and level of economic standards) with volunteer Big Siblings. Additionally, there existed a waiting list of over 200 youth anxious to be matched with a Big Brother or Sister. For the analysis a random stratified sampling strategy was used to select a comparison group from the matched and non-matched groups. The questionnaires were administered year round, thus controlling for any seasonal influences affecting the results.

RESULTS

A t-test analysis was used to compare the scores of the matched and not-matched youth. The results clearly indicate that the matched youth scored significantly higher on the questionnaire than did the non-matched group of youth as shown in Table 3.

TABLE 3
MEANS, STANDARD DEVIATIONS AND OVERALL t-TEST FOR DIFFERENCES BETWEEN MATCHED AND UNMATCHED GROUPS

	Not Matched	Matched
Mean	3.532	3.818
Standard Deviation	0.479	0.394

t-value = 8.95 ($p < 0.001$).

While the evidence for the program's effectiveness appears strong, it is possible that the obtained scores are not as noteworthy as they seem. There is the possibility that no significant difference exists between group scores for all but a few questions. In other words, if the questions were sorted by type or program area, one might find that the mean differences between the two groups varied little except for one or two question areas. Such a pattern of results would lead to a different conclusion about the program than if the results were consistently positive and statistically significant across a wide spectrum of activities and types of questions.

Thus, group differences for the three types of questions and four program areas were also examined. The results of the t-test were positively significant ($p < 0.001$) for the matched youth on all three types of questions (chance, feeling and attitude). This is good news for the program staff, who claim to be positively affecting youth in all three of these domains.

Next, the analysis compared the results among the four program areas (Group Activity Skills, Individual Activity Skills, Group Social Skills, and Individual Social Skills). Figure 2 helps illustrate the pattern of results, and was constructed by sorting each child's questionnaire first by program area and then by type of question. The mean for each program area was calculated and subtracted from the mean of the matched and non-matched groups. Thus, a positive score indicates a group score above the mean; a negative score indicates a group score below the mean.

Two important points are evident from this graph. First, the direction of the differences indicates that the matched youth consistently score higher on the questionnaire across all the program areas. This provides positive feedback for the One-to-One staff, indicating that they are providing services which satisfy a variety of needs. Second, the pattern of results (specifically that the matched youth scored highest in the Individual Social Skills program area) meets the pre-data collec-

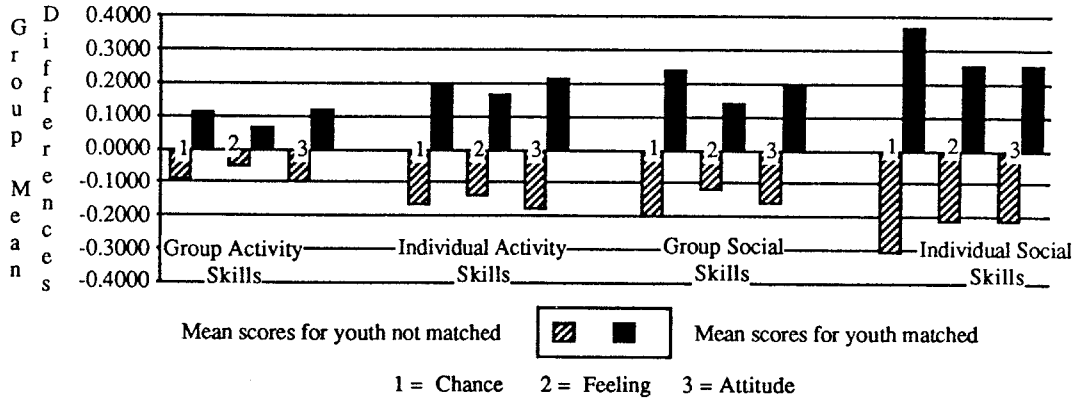


Figure 2. Differences between matched and unmatched groups for the four program areas.

tion predictions of the staff about the program. Since the emphasis of the One-to-One staff training as well as supervision of volunteers is in this area, the staff expected the questionnaire results to reveal a stronger difference between the groups of youth in this program area as compared to other areas. The reasoning for this expectation is as follows: Most youth have other programs available to them which encourage participation in physical activities, but many youth do not have access to the leadership in the personal development area.

The *t*-tests for significance of these results are not as powerful compared to earlier results. This is to be expected since the number of items per test were reduced as the data were further subdivided. Note, however, in Table 4 that the results are significant in the Individual Social Skills area, as predicted by the One-to-One staff.

TABLE 4
MEANS, STANDARD DEVIATIONS AND *t*-TESTS FOR DIFFERENCES BETWEEN MATCHED AND UNMATCHED GROUPS FOR THE FOUR PROGRAM AREAS

	Not Matched	Matched	*t*-value	*p*-value
Group Activity Skills				
Mean	3.540	3.747	1.06	.291
SD	0.607	0.648		
Individual Activity Skills				
Mean	4.060	3.703	2.36	.022
SD	0.514	0.566		
Group Social Skills				
Mean	3.280	3.639	1.81	.076
SD	0.813	0.717		
Individual Social Skills				
Mean	3.278	3.318	2.51	.015
SD	0.760	1.008		

CONCLUSIONS

These evaluation results provide evidence for the effectiveness of the One-to-One program. Concept mapping provided an invaluable tool for conceptualizing the program, designing a questionnaire, strengthening the validity of the results and enabling predictions of the results. Finally, concept mapping proved to be a manageable method, requiring only a few hours of staff time.

Evaluation and Program Planning, Vol. 12, pp. 59-65, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

IN THE EYE OF THE BEHOLDER: RELATIONAL AND HIERARCHICAL STRUCTURES IN CONCEPTUALIZATION

LESLIE J. COOKSY

Cornell University

ABSTRACT

This paper describes the translation from hierarchical structures, operationalized as outlines, to relational structures, operationalized as the maps derived from a multidimensional scaling and cluster analysis procedure. Using a variety of models that represented different assumptions about the relationships within the outline, similarity matrices were derived for a sample outline. The resulting maps were examined to see how well they reflected the outline. The three best reflections are presented here. The ability to translate between hierarchical and relational structures will give program planners and evaluators greater flexibility in both the sources for and the representation of a conceptualization.

Information about interrelationships between ideas or objects is usually presented in either hierarchical or relational form (Novak & Gowin, 1984). Outlines and narratives are examples of hierarchical forms. Maps and correlational tables are common forms of relational information. Both hierarchical and relational representations of information have advantages and disadvantages. For example, outlines and narratives can be more detailed than maps without losing clarity, while the information on maps can present a lot of information without the use of a lot of words. This facilitates the access to the information by the beholder of the map and also gives him/her more responsibility for the interpretation of the information.

The purpose of this paper is to explore the possibility of developing models which allow us to translate from one form to another. The ability to generate hierarchical structures from concept maps has already been demonstrated by Trochim and Linton (1986). They derived an outline from a concept map by first grouping the entities in a cluster tree based on a hierarchical cluster analysis. Then, subheadings and headings for the groups of entities were obtained from the names of map clusters and regions. These names had been determined by the respondents in the process of interpreting the concept map. Since Trochim and Linton (1986) have already described the transition from a relational form to a hierarchical form, the central focus of this paper is the translation of hierarchical information to a relational representation.

The ability to make this kind of transition would be valuable for many reasons. At a theoretical level, it is an interesting question in itself. Researchers have already linked the two different forms of representation to different cognitive styles and perhaps to the dominance of one brain hemisphere over the other (Rico, 1983).

Related to its theoretical value is the suggestive value of developing models for moving from one representational form to another. By viewing the same problem using different structures people may see the issues differently. For example, data from field notes may be organized in a hierarchical (e.g., outline) fashion, but before writing a paper one could display the data in a relational form (e.g., a concept map) as a heuristic device to suggest different ways of organizing the paper. The translation model could free the data from the hierarchy and allows the researcher to view it in new ways.

Another reason for pursuing this research is that by developing models of this type we may be able to utilize the output of computerized outlining programs directly to produce pictorial representations. In the development of concept maps, outlining, computer assisted or not, could be used as an alternative to unstructured sorting.

Requests for reprints should be sent to Leslie Cooksy, 2144 California Street NW, #602, Washington DC 20008.

Finally, this research can be useful in planning and evaluation. Most programs are based, at least in part, on a hierarchy of goals, objectives, and tasks. This hierarchy is represented in organizational charts, reports, and other program documents. Using these documents as input in hierarchical form, a computer program could be devised to automatically translate that hierarchical information into relational or map form. In program planning, the hierarchical information gathered from program documents and displayed in relational form can be used to rethink programs. Information may be more easily synthesized in a relational form than in the form of a written text, and it is also more easily discussed and reorganized. Once the relationships displayed have been discussed, new information could be added to the system, and a new hierarchical representation could be generated.

In the evaluation of a program, maps derived from the hierarchical information in program documents could be checked for validity against maps generated from relational information given by the staff or clients about some aspect of the program. The ability to move from hierarchical to relational forms of representation would also facilitate the pictorial representation of evaluation results.

DEVELOPING THE MODELS

While the ultimate goal of this research is to develop models for moving easily from one form to the other, this paper focuses on the translation of hierarchical information to relational structures. Hierarchical structures were operationalized as outlines because of the simplicity and utility of outlines in the reduction of more complex hierarchical data. Relational structures were operationalized as the maps that are derived from the multidimensional scaling and cluster analysis techniques described in the first paper in this volume.

In order to find a model that would enable us to develop models to generate a map that was a fair representation of an outline, the following procedures were used. First the relationships between entities in an outline were established according to several major assumptions. Then similarity matrices based on each set of assumptions were derived. Finally, multidimensional scaling and cluster analysis techniques were used to generate the maps based on the matrices. To illustrate the models we explored, an outline of "Threats to Validity" was used. This outline, derived from the familiar taxonomy described by Cook and Campbell (1979), is shown in Figure 1. Three sets of models are described here. While the models differ in their assumptions, each uses the same counting rules to arrive at the similarity matrix. These rules are:

1. In each model, the line that directly connects two entities without passing through a third entity equals one path.
2. The distance between any two entities in the outline is equal to the number of paths between them.
3. The similarity between any pair of entities is equal to the number of paths between them subtracted from the largest number of paths possible between any two entities.

Each of the three models are considered in order below.

Model 1: Hierarchical Model

The first model is depicted graphically in Figure 2. To see how this model works, consider how the counting rules were applied to generate a similarity matrix for the elements in the outline according to this model. An example of a path is the line between "Internal Valid-

THREATS TO VALIDITY

I. Internal Validity

A. THREATS TO CONCLUSION VALIDITY

1. Low Power
2. Violated Assumptions
3. Unreliable measures

B. THREATS TO INTERNAL VALIDITY

1. Statistical Regression
2. Selection
3. History

II. External Validity

A. THREATS TO CONSTRUCT VALIDITY

1. Inadequate preoperationalization
2. Mono-operation bias
3. Mono-method bias

B. THREATS TO EXTERNAL VALIDITY

1. Interaction of selection and treatment
2. Interaction of setting and treatment
3. Interaction of history and treatment

Figure 1. Outline used to test models.

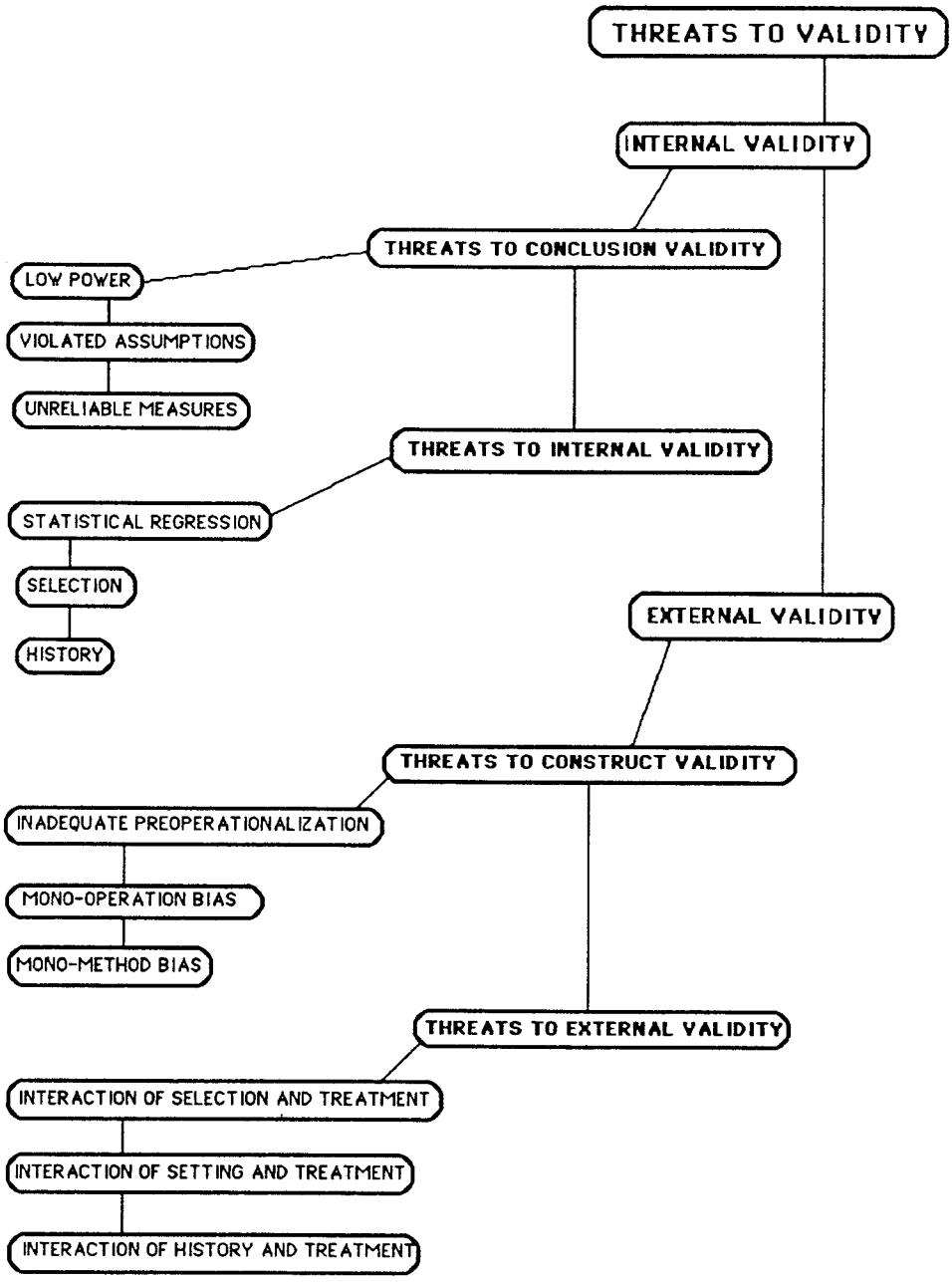


Figure 2. Hierarchical model.

ity" and "Threats to Conclusion Validity." Thus, the distance between "Internal Validity" and "Statistical Regression" is equal to three paths. The similarity of "Internal Validity" and "Statistical Regression" is equal to the greatest possible distance between any two entities in the model, which is twelve, minus the distance between these two, which is three, so the similarity is equal to nine. The model in Figure 2 is a simple hierarchy which assumes that the concept "Internal Validity" is closer than "External Validity" to "Threats in the Validity;" that "Mono-operation Bias" is closer than "Mono-method Bias" to "Threats to Construct Validity"; and so on. In more general terms, the model assumes that the first subheading is more related to its major heading than the second, and so on.

The map based on this model derived from the similarity matrix is shown in Figure 3. The map seems to be a fair reflection of the rather strict hierarchical assumptions made. However, one might have expected "Threats to Validity" to be to the left of "Internal

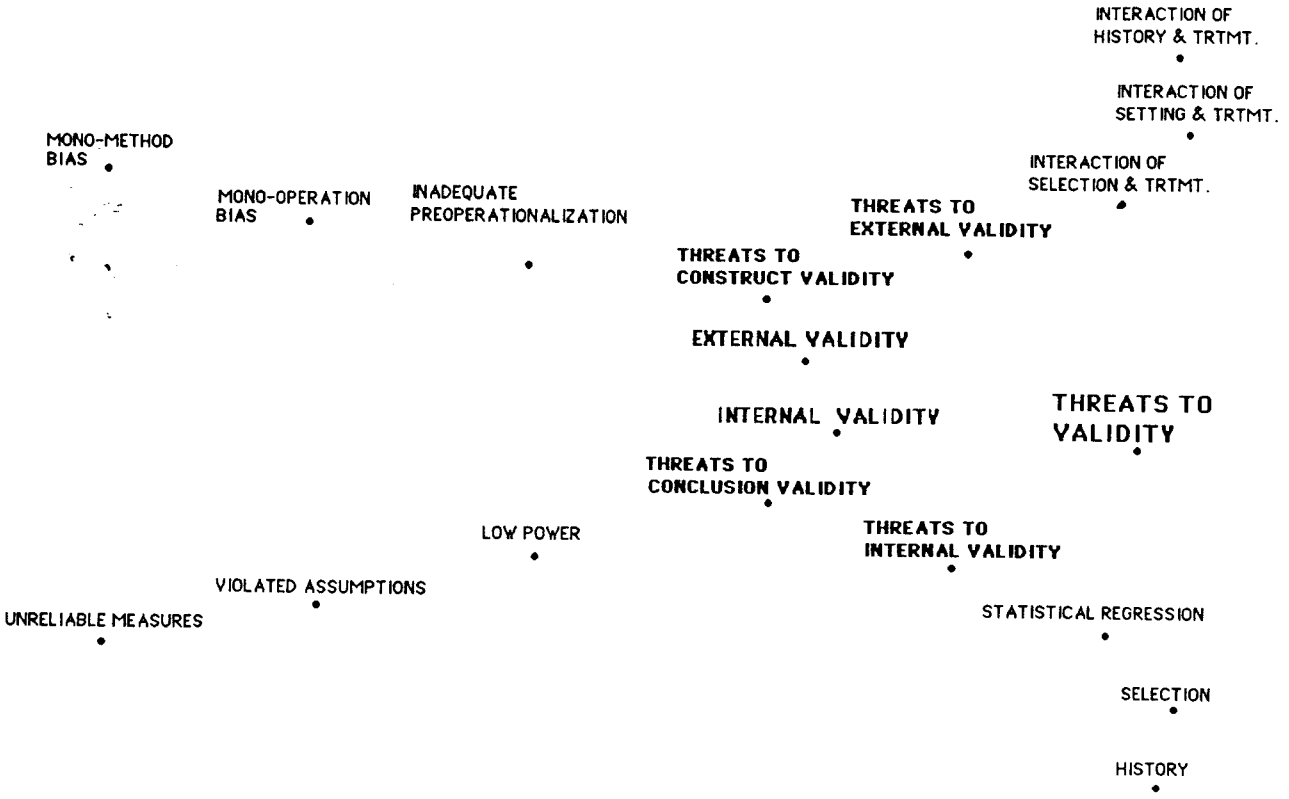


Figure 3. Map based on hierarchial model.

Validity" and "External Validity" since this would place it closer to "Threats to Conclusion Validity" and "Threats to Construct Validity" as it is in the model. Not only is the placement of the main topic not completely congruent with the model, it may not be the best reflection of the outline. One can think of the main topic as being "central" to the outline. It should therefore be in the center of the map.

Model 2: Digraph Model

The second model, shown in Figure 4, addresses this issue. The assumptions of this model are based on Waller's (1976) discussion of directed graphs, or digraphs. Unlike the first, totally hierarchical model, the second model displays subheadings under a heading as equally distant from all other headings. Thus, the main topic is centrally related to the subdivisions. For example, in this model "Internal Validity" and "External Validity" are considered equally distant from the overall heading of "Threats to Validity." As a result, the map derived from this model should have "Threats to Validity" in the center. An additional advantage is that, because of the symmetry of this model, much less information than that used in the hierarchical model is required to derive the similarity matrix.

The map derived using the second model can be seen in Figure 5. While the map based on this digraph model has the main topic in a central place and otherwise sensible placement of the entities, it is less useful than the map based on the hierarchical model simply because the entities at the most specific level (e.g., I.A. 1-3) are not differentiated. For this reason, the next map was based on a set of assumptions that combined the digraph and the hierarchical models.

Model 3: Hybrid Model

The third model, combining the hierarchical and digraph models, is shown in Figure 6 and the resulting map in Figure 7. This model is identical to Model 2 except that at the lowest subheading level the subheadings are considered hierarchical as in Model 1. This last restriction will force the lowest subheading entities to be differentiated on the map. The resulting map is a better representation of the outline than that obtained from the other two models. Each entity is placed separately and symmetrically on the map in a way that reflects assumptions made about the structure of the outline.

In addition to the maps and models displayed here, other models were attempted using even less information than that given in the digraph model. However, the results were not especially useful and those attempts were abandoned.

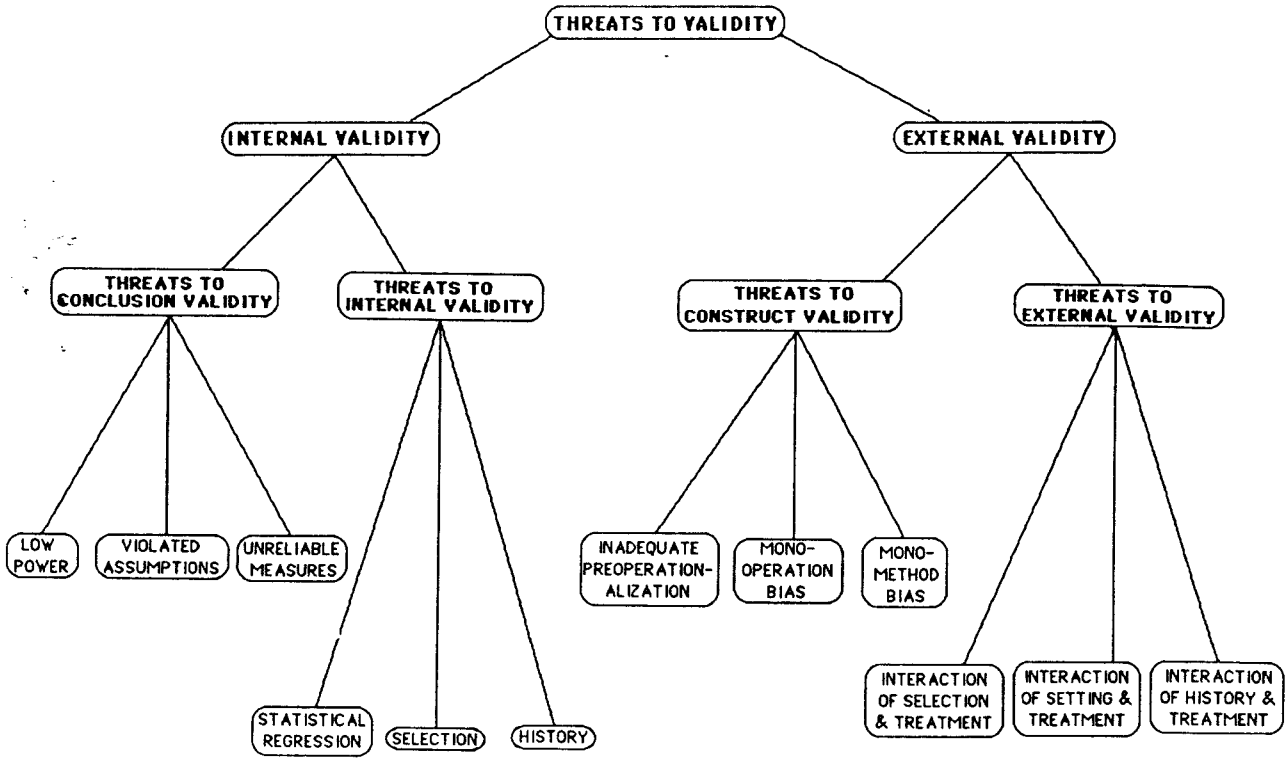


Figure 4. Digraph model.

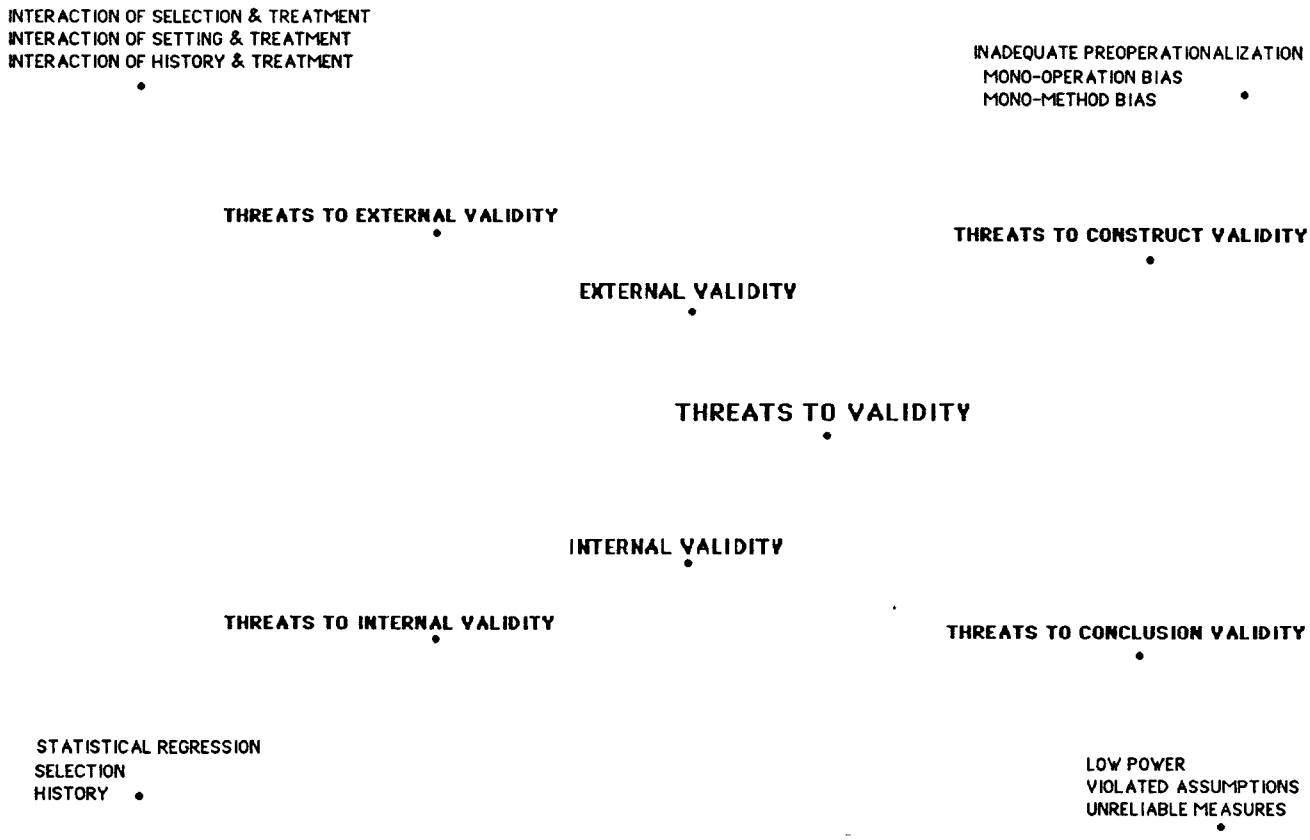


Figure 5. Map based on digraph model.

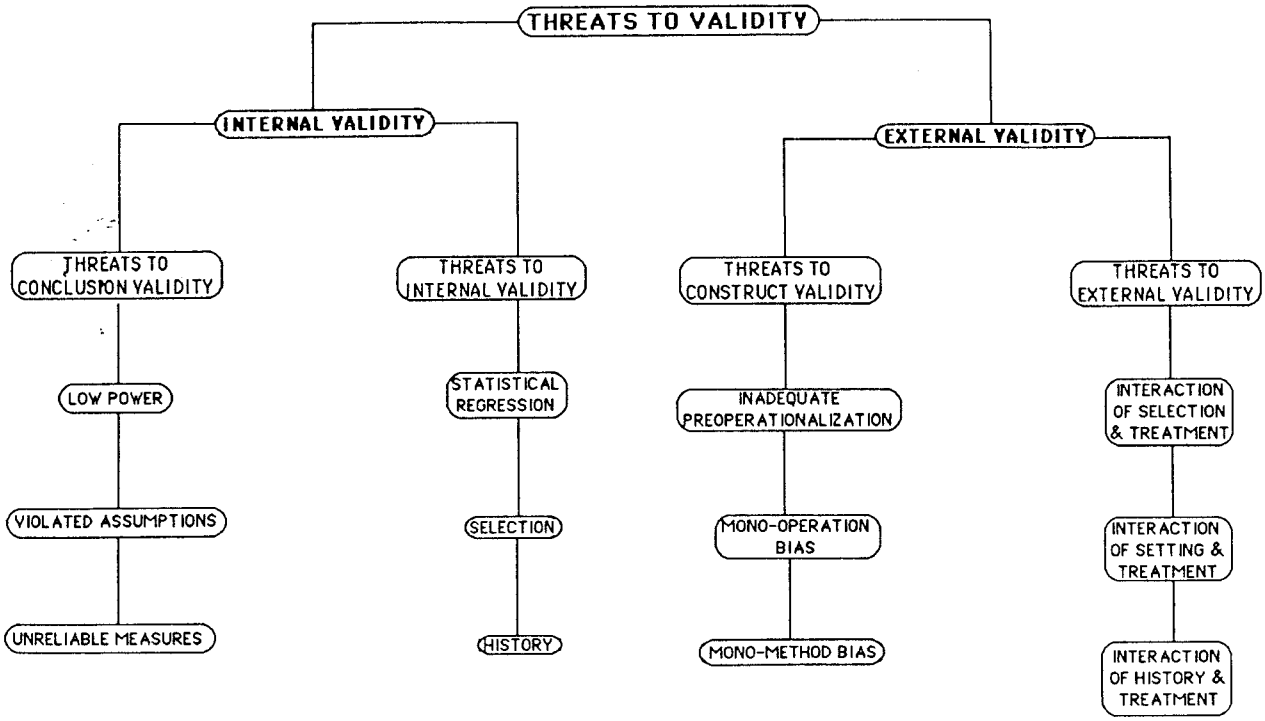


Figure 6. Hybrid model.

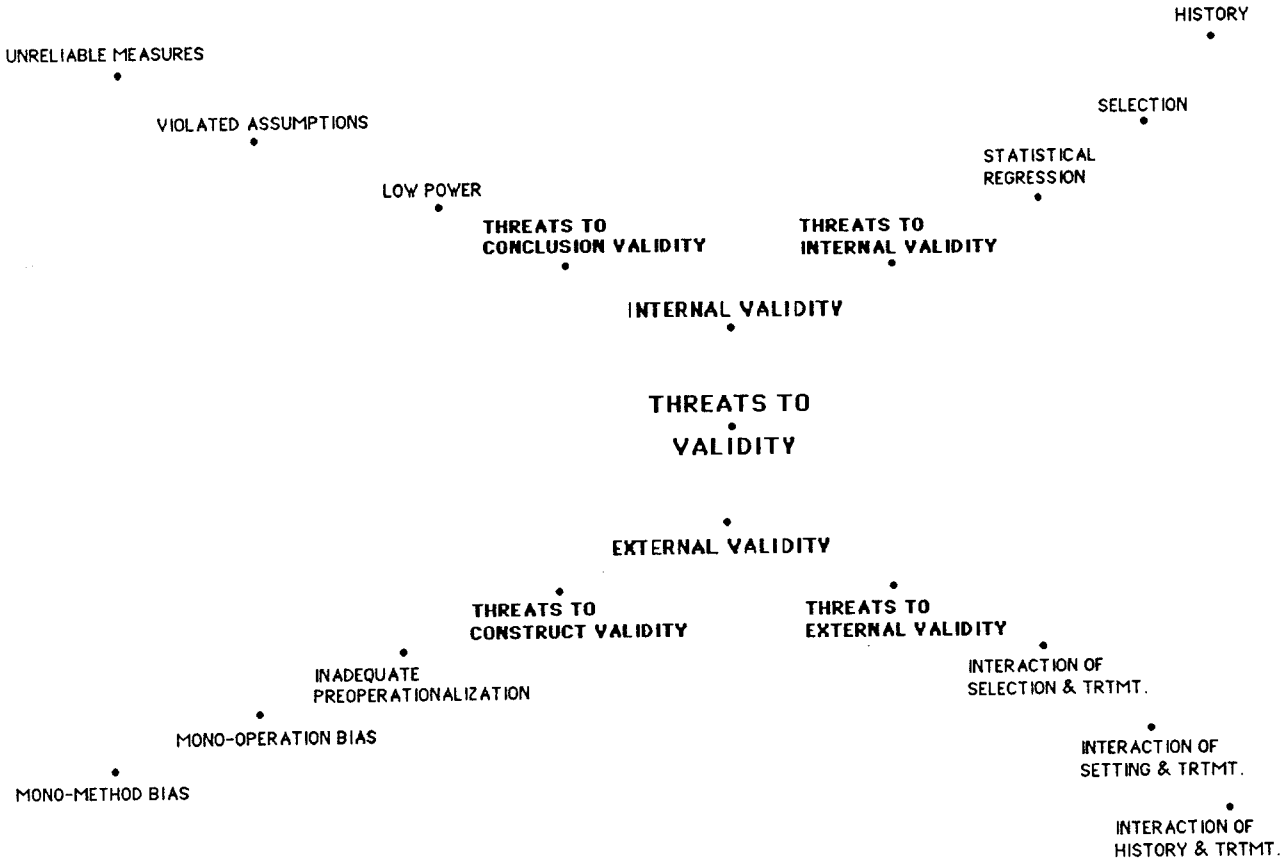


Figure 7. Map based on hybrid model.

CONCLUSIONS

Our general conclusion is that the best model is one which combines both the hierarchical (Model 1) and digraph (Model 2) approaches. The digraph approach is used overall, and only the last nested group under each branch of the outline is broken down hierarchically. This is done so that these lowest-level entities will be distinguished from each other on the map.

While this effort is just the beginning of a wider attempt to develop models to translate automatically from one form of representation to another, it is quite promising. In the future, techniques like this may be used to help program planners and evaluators involve stakeholder groups in developing a unified conceptualization of the program. Each person would use a common set of entities and place them in outline form to reflect their view of superordinate-subordinate relationships within a set of entities. The similarity matrices derived from the outlines could then be aggregated, and a single map would be generated. This map would be a single representation of the participants' conceptualizations of the program. New information from discussions of the map could be added and then translated back into outline form for the writing of a report.

Future work in the development of this technique will include the programming of the digraph-hierarchical model for use with the multidimensional scaling and cluster analysis procedure and the application and testing of the method with more complex and realistic hierarchical structures.

REFERENCES

COOK, T.D., & CAMPBELL, D.T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago, IL: Rand McNally College Publishing Company.

NOVAK, J.D., & GOWIN, D.B. (1984). *Learning how to learn*. Cambridge, England: Cambridge University Press.

RICO, G.L. (1983). *Writing the natural way: Using right-brain techniques to release your expressive powers*. Low Angeles, CA: J.P. Tarcher.

TROCHIM, W.M.K., & LINTON, R. (1986). Conceptualization for planning and evaluation. *Evaluation and Program Planning*, **9**, 289–308.

WALLER, R.J. (1976). The synthesis of hierarchical structures: Technique and applications. *Decision Sciences*, **7**(4), 659–674.

Evaluation and Program Planning, Vol. 12, pp. 67–74, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

USING CONCEPT MAPPING FOR PLANNING THE IMPLEMENTATION OF A SOCIAL TECHNOLOGY

MARC MANNES

Cornell University

ABSTRACT

This research utilized concept mapping to analyze and compare the process of planning to implement a specific social technology—known as family based intensive in-home services (FABIIHS)—which has demonstrated effectiveness in certain settings and is being adopted in two new sites (the Upper Peninsula of Michigan and the Rio Grande Corridor in New Mexico). In each site, planning teams were formed to identify the factors which needed to be dealt with in order to implement a FABIIHS project in their setting. Planning team members consisted of direct service agency workers at the managerial and case practice levels from Indian agencies, state, and non-profit organizations which were part of the community service network and which therefore had to be involved to successfully implement the social technology of FABIIHS. Approximately 35 people participated in the three-step process which involved: (a) the generation of a set of eighty-one implementation factors; (b) sorting the factors and rating each one for their importance and ability to influence; and (c) interpretation and naming of the concept map. On the basis of this project, concept mapping is seen as a potentially valuable process which can be helpful in planning for the implementation of new social technologies.

To date there had been insufficient research on the process of planning to implement innovations in the human services. These innovations, more appropriately titled social technologies, are novel and technical means by which the human services accomplish their objectives (Thomas, 1978). They result from the fusion of research and development methodologies transferred from engineering and industry to the human services with experiential knowledge known to many as "practice wisdom" (Whittaker & Pecora, 1981). The development of social technologies is typically supported through research and demonstration grants provided to public agencies, non-profits, and institutions of higher education. Research on social technologies has primarily been evaluative in nature and as a result has emphasized outcomes (post-implementation results), while tending to ignore the planning process (pre-implementation and implementation issues and factors).

The exploratory research described in this article utilized concept mapping to analyze and compare the process of planning to implement a specific social technology which has demonstrated effectiveness in certain settings and is being adopted in two new sites. Specifically, concept mapping was used to study and compare the perceptions of human service staff employed in direct service agencies in two distinct sites who are planning to implement a social technology known as family based intensive in-home services (FABIIHS).

The stimulus for the development of social technologies like FABIIHS can be traced to the Adoption Assistance and Child Welfare Act of 1980, Public Law 96-272, which was passed to alter the way in which the public child welfare system was serving dependent children suffering from abuse and neglect. The law was designed to support and preserve the integrity of families, reduce the number of children "stranded" in the foster care system, set guidelines for permanency planning, and reverse federal financial incentives which had made foster care placement an immediate and seemingly advantageous choice when deciding how to respond to abuse and neglect cases. The law sought to keep families intact by preventing the unnecessary sep-

Requests for reprints should be sent to Marc Mannes, 604 Valencia Drive N.E., Albuquerque, NM 87108.

aration of children from their parents, and emphasized the importance of providing services to support and strengthen families in an attempt to avoid removing the child(ren) and placing him or her in foster care.

In the attempt to preserve and strengthen families and reduce the number of children receiving substitute care, the federal Department of Health and Human Services' Office of Human Development Services in its R&D Coordinated Discretionary Funds Program for FY87 solicited proposals to administer demonstration projects addressing the theme of Foster Care Placement Prevention. A major innovative social technology in this area can be labeled family based intensive in-home services (FABIIHS). The social technology, as its name implies, consists of a family-focused intervention, of an intense and time-limited nature, delivered to clients in their own home. The objective is to quickly attempt to alter dysfunctional family behavioral and attitudinal patterns so children can remain with their biological parents.

The American Indian Law Center, Inc. (AILC), a national organization offering broad-based training and technical assistance to tribal and other governments in the areas of policy analysis, intergovernmental relations, planning and evaluation, and law-related services, was funded to administer a placement prevention effort entitled Project INTACT. Project INTACT sought to work with tribal human service agencies to promote and facilitate the development of placement prevention models consistent with FABIIHS characteristics, and assess the planning process to implement such programs. Concept mapping was seen as a potentially valuable tool for planning Project INTACT implementation.

METHOD

Sample

This project involved activity in two distinct implementation sites–The Upper Peninsula of Michigan and the Rio Grande Corridor in New Mexico. Planning teams were formed in each site to identify the factors which needed to be dealt with in order to implement a FABIIHS project in their setting. Planning team members consisted of direct service agency workers at the managerial and case practice levels from tribal governments, Indian agencies, state, and non-profit organizations which were part of the community service network and which therefore had to be involved to successfully implement the social technology of FABIIHS.

Developing the Concept Map

An initial series of planning meetings was conducted. Planning team members in each site were assembled and provided with a half-day presentation on family based intensive in-home services so there would be a common understanding of the structure of such programs, their characteristics, the nature of services, etc.

The first step in the concept mapping process was to have the members of both planning teams generate statements, or in this case what may be more appropriately labeled "implementation factors"—the items which would need to be dealt with to make a FABIIHS project operational. The brainstorming prompt used to generate these implementation factors is consistent with Williams' (1975) argument that implementation needs to focus concretely on what needs to happen to get something, in this case a social technology, put into practice. It was also assumed the relatively unstructured nature of the brainstorming activity would allow for both "technical" and "political" planning factors to be cited. The prompt was also structured so respondents would consider not just what they needed to do, but would also be motivated to think in a broader sense by considering what needs to happen in general. The specific prompt was:

In order to get a family based service project operating:

- What do you need to do?
- What do you think needs to be done?
- In other words, what needs to happen?

Twelve persons participated in the first New Mexico Planning Team session; twenty people were involved in the Michigan Planning Team session. The Michigan team generated ninety-one (91) implementation factors while the New Mexico team developed a list of fifty-three (53) factors.

In order to engage in a comparative analysis of implementation sites a master list of implementation factors needed to be devised. It would have been ideal to have both teams review each other's list and have them come up with a unified master list. Unfortunately, time and funding constraints precluded this option. Instead, two staff members of the American Indian Law Center's Project INTACT reviewed both lists to identify redundancies, and develop a master list which accurately reflected the range of items produced by the members of both planning teams. The master list contained eighty-one (81) implementation factors and is shown in Table 1. Each of the eighty-one (81) implementation factors was numbered and printed on a separate slip of paper.

In the second step of the process, all planning team members were asked to sort the master list of implementation factors into piles in a "way which makes sense to you." This was done to obtain a conceptual portrait of how the implementation factors were orga-

TABLE 1
LISTING OF EIGHTY-ONE STATEMENTS FOR CONCEPT MAPPING

1. Educate the community to understand the need for family based services
2. Define who the family is
3. Identify funding sources
4. Establish qualifications for providing services
5. Provide training on family based services
6. Assess current caseload for service needs
8. Have funding sources support tribal innovations
9. Funding agencies need to do less defending of turf
10. Define program from the tribal perspective, not solely from the funding source's perspective
11. Consider staffing needs
12. Reduce overlapping and/or conflicting roles of the workers
13. Identify families eligible for services
14. Educate and gain support of tribal council
15. Get tribal council to set family based services as a top priority
16. Structure services within tribal social services agencies
17. Advertise, recruit and hire workers
18. Consider who and what type of workers are needed
19. Teach the community about family based services
20. Get funding source to provide child welfare dollars to support family based services
21. Determine appropriate services for the family based services program
22. Secure funds (grantsmanship)
23. Establish job descriptions
24. Distinguish between chronic and "new" crisis families
25. Gain support of other service providers
26. Providers need to inform community of availability or their interest in providing family based services
27. Provide training in counseling
28. Obtain community support
29. Obtain data on number and types of current placements
30. Consider worker liability issues
31. Establish inter-agency agreements
32. Get funds for training
33. Set program goals and objectives
34. Provide concrete services, e.g., transportation
36. Coordinate for effective referrals
37. Be willing to terminate unsuccessful cases
38. Distinguish between prevention of abuse and neglect, and prevention of placement
39. Train supervisors jointly with workers
40. Develop short position paper on Indian family based services for tribes and Indian organizations
41. Recruit committed and compassionate staff
42. Define service catchment area
43. Examine impact of changing services on existing agencies
44. Network with courts
45. Establish data base of families served
46. Provide limited caseload for workers
47. Have program manager who is a skilled social worker
48. Provide on-going education and training
49. Set realistic family goals
50. Conduct public relations efforts
51. Catalog existing resources
52. Establish strong linkage to aftercare
53. Finalize Indian child welfare policy(ies)
54. Upgrade workers' salaries
55. Organize inter-tribal political efforts
56. Work with families in an environment where they will be receptive
57. Prevent high staff turnover
58. Develop service standards
59. Reduce jealousies and turf issues
60. Retain a flexible clinical approach
61. Collect information on other programs and adapt for own model
62. Prevent worker burnout
63. Encourage universities to teach family based services
64. Generate program operations guidelines
65. Provide worker incentive program
66. Establish inter-agency coordinating council to identify, motivate, foster collaborative relations
67. Coordinate with other providers to promote long term goal of family unity
68. Develop research and evaluation plans
69. Computerize communication systems
70. Focus on family unit
71. Learn about culture and incorporate into program
72. Establish appropriate supervision
73. Have leaders model appropriate behavior
74. Identify potential problems and barriers
75. Choose appropriate approach, e.g., team or individuals
76. Train non-Indian community to work effectively with Indian clients and agencies
77. Workers should model behavior for families
78. Address creaming effects, i.e., "easier" cases are accepted, or "hardest" cases are referred
79. Strengthen existing programs before developing new [ones]
80. Upgrade housing
81. Establish service priorities

nized and interrelated in the minds of planning team members.

In order to analyze and interpret the results of the sorting activity, the data from the sorts were entered into the Concept System software package developed by Trochim (this volume). Initially a multidimensional scaling (MDS) analysis was performed to scale the 81 factors in relation to one another based on item rating similarities. A two-dimensional MDS solution was employed to ease interpretation and comprehension. This allowed for the production of three distinct maps representing the geometric configuration of the 81 factors. The first was a composite map representing both planning teams. The second map represented only the Michigan team, and the third depicted the data for only the New Mexico team. The maps from each of the sites appeared similar and, consequently, the composite map was used. This map is shown in Figure 1. The statements on the map were grouped using Ward's algorithm for hierarchical cluster analysis. A 25 cluster solution was chosen as the starting point, and then fewer-cluster solutions (i.e., 20, 15, 6) were also generated. The various cluster solutions were subjectively and intuitively examined by Project INTACT staff who

Figure 1. MDS point map for Project INTACT.

used their knowledge and experience to determine the level at which both significant differentiation and meaningful categories of similar factors occurred. Combining the hierarchical cluster analysis with the distinction and commonality criteria, a six cluster solution appeared to be the most appropriate and useful.

From the perspective of Project INTACT staff, the implementation factors in Cluster 1 seemed to focus on engaging in activities which set the stage for FABIIHS to begin. Cluster 2 appeared to deal primarily with interagency matters while Cluster 3 tended to concentrate on what might be considered developmental and operational issues. Cluster 4 centered on funding concerns and Cluster 5 emphasized services. Finally, Cluster 6 targeted what we might call human resource matters. Project staff were anxious to find out how the planning team members themselves would define the clusters.

A second wave of planning meetings was conducted with the planning team representing each implementation site. There were several objectives for these sessions: (a) to have planning team members rate the implementation factors in terms of how important they perceived each implementation factor to be and to what degree they perceived they could influence each factor or make it happen; (b) to review the six cluster solution and see if planning team members saw it as insightful and appropriate for continued use; (c) assuming that the six cluster solution was acceptable, to have members of the teams come up with labels so they would ultimately be the ones to define the various clusters.

For both the importance and influence ratings each implementation factor was rated on a 7 point Likert-type response scale with 1 being almost no importance/influence, 4 being moderate importance/influence, and 7 being extreme importance/influence. Those members of the team who were not present for the second meeting had both instruments mailed to them and were asked to complete and return them to the Law Center; 9 out of 10 people returned completed instruments. A total of 37 people completed both instruments – 18 from New Mexico and 19 from Michigan.

When planning team members completed the instruments, they were asked to consider the six cluster solution. Both planning teams agreed it was a useful means of interpreting and making sense of the implementation factors. Once the six cluster solution was accepted, planning team members in attendance were asked to come up with a name for each cluster. Unlike the instruments gauging perceptions of importance and influence, which were mailed to those members of the planning team not attending the second meeting, only those persons attending the second meeting at each implementation site were asked to generate names for the six clusters. Moreover, not all people attending the meetings turned in a sheet. As a result, only 23 people responded to either some or all of this portion of the study.

Given sufficient time it would have been ideal to have the planning team members work through the various names they had given to each of the six clusters or implementation domains and reach consensus on a name for all six. Unfortunately there wasn't sufficient time to allow for that to occur so Project INTACT staff made use of a wordcount procedure to try and come up with a name for each domain.

For cluster 1 the words used the most were: community (17 times); education (15 times); tribe/tribal coun-

cil (5 times); and support (4 times). There was an almost even split in the use of these words between the members of the planning team in each implementation site. Three more people in Michigan than in New Mexico use the word community, but there were 3 more respondents from Michigan. Therefore, it seems reasonable to define this implementation domain as COMMUNITY EDUCATION FOR SUPPORT.

For cluster 2 the words used most were: coordination (10 times); networking (7 times); collaboration (4 times); interagency and service providers (both 3 times). Virtually every respondent included at least coordination, collaboration, networking, or cooperation in their label. It is apparent the second implementation domain can be labeled COORDINATION AMONG AGENCIES.

For cluster 3 the word family was used (13 times), services (9 times), identification (8 times), and need (4 times). Again there was an almost even split in use of the words by the members of each team. The third implementation domain can be labeled FAMILY IDENTIFICATION FOR SERVICES.

For cluster 4 funding or fundraising was mentioned (18 times), with another word dealing with financial matters, fiscal, cited (2 times). Clearly, the fourth domain can be christened FUNDING.

For cluster 5, program was mentioned (7 times), services was also used (7 times), implementation (4 times), development (3 times), and planning (2 times). It would seem the fifth domain is about PROGRAM AND SERVICE DEVELOPMENT.

Finally in cluster 6 the words used the most were: staff (13 times); training (10 times). The sixth implementation domain can be coined STAFF TRAINING.

Project INTACT staff reexamined the six cluster solution in light of the labeling exercise to decide if the original six cluster solution still seemed appropriate to use for analysis. The criterion used was to consider if the six cluster solution accomplished adequate differentiation and similarity in the grouping of implementation factors consistent with the results of the labeling effort.

After additional review of the six cluster solution by Project INTACT staff, several of the implementation factors seemed to be grouped in a category which did not make sufficient sense, or seemed totally inappropriate. This is not surprising since the hierarchical cluster analysis only groups according to spatial relationship, and fails to take into account any other criteria. After additional review of the spatial position of several of the implementation factors in the six cluster solution several minor modifications were made to the computed map. These are illustrated in the map shown in Figure 2. The figure shows both the original (broken line) and modified (solid line) six cluster solutions. Overall, six (6) implementation factors were moved to a cluster which made more sense for them to be in.

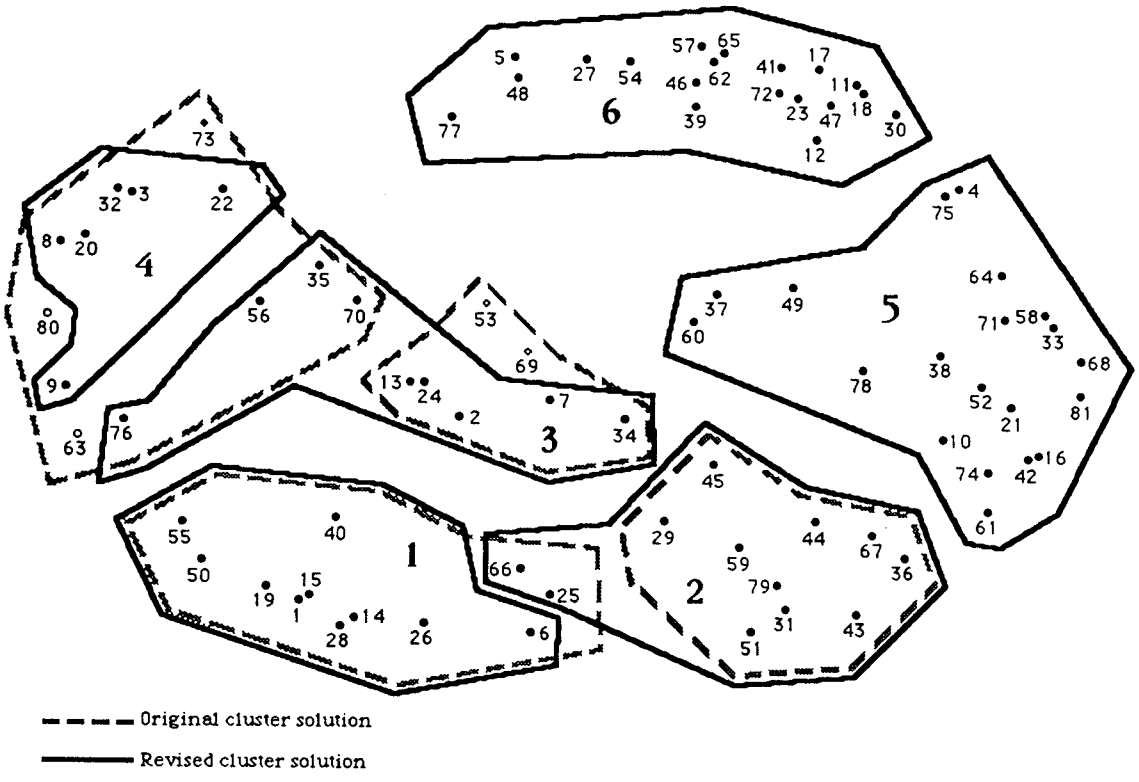


Figure 2. Original and modified six-cluster concept map for Project INTACT.

After some thought and discussion about their connection to the cluster they were placed in, five (5) implementation factors were completely eliminated from additional analysis. Clusters 5 and 6 were not modified at all.

More specifically, if one looks at Figure 2 it is clear that implementation factor #25—and implementation factor #66 are on the border between Clusters 1 and 2 and probably could have been placed in either cluster, although the six cluster solution chose to place both of them in Cluster 1. Project INTACT staff decided they both fit better in Cluster 2. Note that it was decided to keep implementation factor #6 which is spatially very close to numbers 25 and 66 in Cluster 1 and not move it.

The same principle was at work in transferring four other implementation factors—numbers 35, 56, 70, and 76. Again, referring to the concept map in Figure 2 reveals implementation factors 35, 70, and 56 could have just as easily been assigned to Cluster 3 or 4, but were assigned to Cluster 4. Yet, they make more sense in Cluster 3, so they were moved. Implementation factor 76 seemed inappropriate in Cluster 4 and appeared more relevant to Cluster 3 so it was shifted.

Five factors were eliminated from the master list for subsequent analysis. Three of these—63, 73, and 80—had been grouped in Cluster 4 in the original six cluster solution. Yet, none seemed directly related to the dominant theme of that cluster. Thus, there was concern that they were outliers, that their inclusion would bias the analysis of that particular cluster. Instead of somewhat arbitrarily placing them in another cluster they were simply eliminated. Two others eliminated were numbers 53 and 69. They had been placed in Cluster 3, yet upon review by Project INTACT staff they appeared tangential to the focus of the cluster.

This highlights an important issue when using concept mapping. The statistical processes, and in particular the cluster analysis, must be weighed against judgment and experience. Our shifting of certain implementation factors demonstrates the importance of incorporating experience and common sense when utilizing concept mapping.

Analysis of the Importance and Influence Rating Data

Once the cluster names and their factors had been agreed upon, the map was used to display the results of

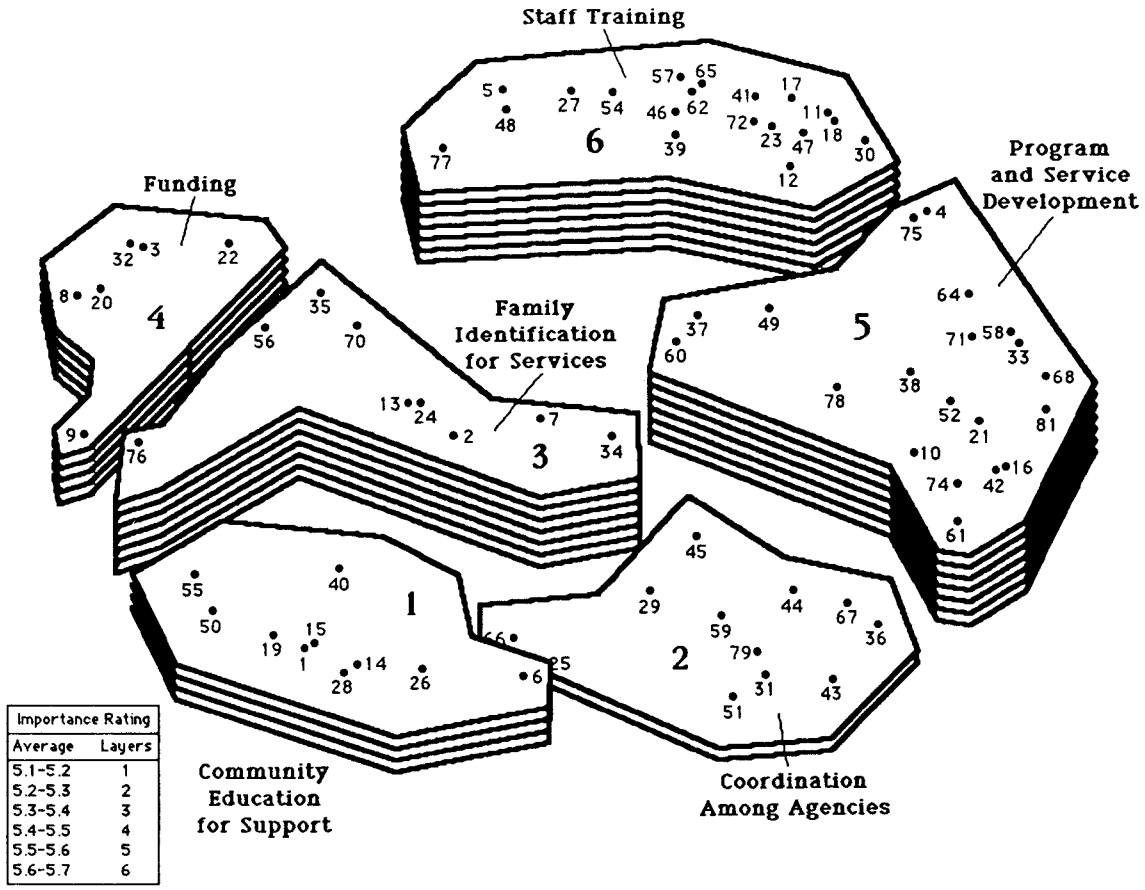


Figure 3. Importance rating concept map for Project INTACT.

the ratings. Here, the results are shown by cluster for the importance ratings in Figure 3 and for the influence data in Figure 4. A comparison of the two is particularly useful. First, examination of the legends on each figure indicates that in general importance ratings were higher than influence ratings. This is not surprising—in many contexts people feel relatively powerless to influence factors which they consider important. Second, both maps show a fair amount of agreement—in both cases, participants attached relatively high importance and ability to influence to the clusters Staff Training, Program and Service Development, Family Identification for Services, and Community Education for Support. Finally, it is perhaps most revealing to note the two major clusters for which the ratings diverged—Funding was seen as important but not easily influenced, whereas Coordination Among Agencies was seen as relatively unimportant but relatively easy to influence. Clearly, it is worth exploring with the participants what the implications of these ratings are for the implementation of a FABIIHS project.

CONCLUSIONS

The use of concept mapping to engage in a comparative analysis of planning to implement the social technology of FABIIHS has proven to be extremely insightful. It helped to generate the implementation factors which need to be dealt with in order to operationalize a FABIIHS project, as seen from the critical perspective of the staff who work in direct service agencies in two distinct implementation sites. Moreover, the process has resulted in the specification and labeling of major domains of implementation factors which must be considered when implementing a FABIIHS project in those sites. The rating of the implementation factors in terms of perceived importance and ability to influence documented which factors and domains are most important and most susceptible to influence by direct service agency staff. The analysis of the data contributes to staff at both sites' understanding of what needs to be done and dealt with in order to get the social technology going.

The data generated also lends itself to additional vital

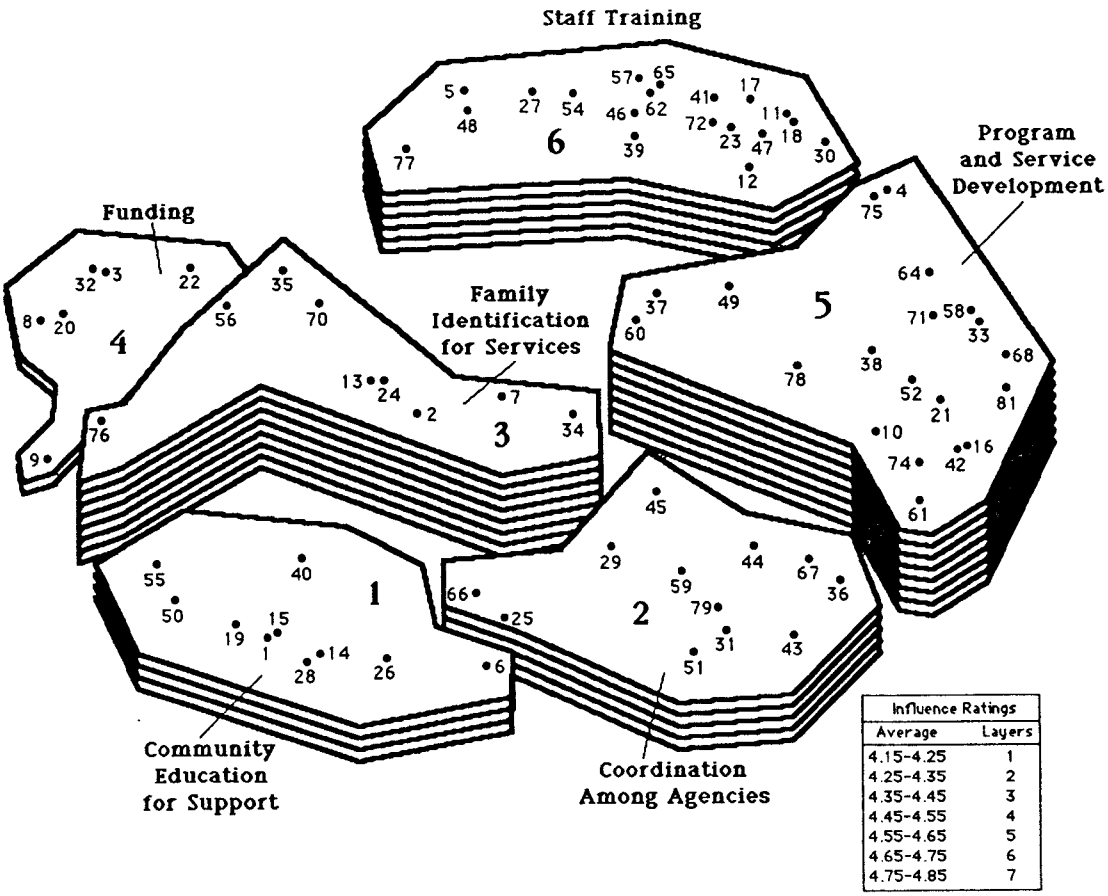


Figure 4. Influence rating concept map for Project INTACT.

analysis to facilitate pre-implementation planning. For one thing, looking at the interaction between importance and influence for each factor and for each domain might engender some useful insights and suggest some important implications for the planning process. Then too, the highest and lowest rated implementation factors in terms of perceived importance and ability to influence, within each domain, could be compared across sites to determine the site-specific or potential generalizability of the findings.

In terms of making broad practical use of the results, these preliminary findings can be used by policymakers and/or managers and direct service staff if they are considering initiating a FABIIHS project in one or more additional sites. From a theoretical standpoint the findings of this study can contribute to moving beyond the global acknowledgment of a "political" and "technical" dimension to planning, to actually describing specific political and technical implementation factors. Also, the findings can begin to define the contours of numerous major domains of implementation factors. By repeating this process with the same, and other social technologies, our understanding of the broad planning for implementation process could be increased. Hypotheses suitable to subsequent study could be created. Finally, in addition to the comparison of perceptions by site, the perceptions of managers could be compared with those of direct service workers, and the perceptions of men could be compared with those of women to see how they differ and to promote theory building.

Very little attention to or study of pre-implementation planning has taken place. This is unfortunate, especially since millions upon millions of dollars are spent by governments on a yearly basis to support the implementation and replication of innovative social technologies to ameliorate social problems. By engaging in a little "front-end" work considerable time and money can possibly be saved so effective social technologies can be implemented in a quicker and smoother manner. Concept mapping is a sophisticated, yet easy to use tool, which helps portray and assess key aspects of the planning process. The use of concept mapping can make a significant contribution to the theory and practice of pre-implementation planning.

REFERENCES

THOMAS, E.J. (1978). Generating innovations in social work: The paradigm of developmental research. *Journal of Social Service Research, 2*, 95-115.

WHITTAKER, J.K., & PECORA, P.J. (1981). The social "R&D" paradigm in child and youth services: Building knowledge convivially. *Children and Youth Services Review, 3*, 305-317.

WILLIAMS, W. (1975). *Implementation analysis and assessment*. Public Policy paper, Institute of Governmental Research, University of Washington.

Evaluation and Program Planning, Vol. 12, pp. 75-80, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

REFINING CONCEPT MAPS: METHODOLOGICAL ISSUES AND AN EXAMPLE

DOUGLAS KEITH

Cornell University

ABSTRACT

This paper describes a concept mapping exercise conducted with a group of public school teachers for the purpose of formulating a plan to integrate computer technology into a conventional educational setting. The exercise involved the generation of conceptual entities through a structured brainstorming session, and a session in which participants subjectively rated the brainstormed entities according to perceived similarities between them. This was followed by a multidimensional scaling and hierarchical cluster analysis of the similarity ratings, and interpretation of the map.

Initial results indicated that there was a confounding effect on the final map because of perceived semantic groupings based on surface characteristics which were not germane to the intended focus of the conceptualization exercise. The paper describes a method that was used to control the effects of these strong, but experimentally irrelevant, category-based associations which subjects in this study made among certain items in a subset of the data. The initial map is presented along with the final map which shows the effect of the attempt to correct for this methodological artifact.

This paper has two purposes. First, it provides an example of the use of the concept mapping procedures outlined by Trochim (this volume) for examining the integration of computers into the school curriculum. In the process of conducting the concept mapping, however, it became apparent that because of the way the brainstorming was handled, the major topic of interest became obscured in the final map. Specifically, the intent of this project was to examine how computers could be integrated into the educational environment. Two separate brainstorming sessions were conducted—one focused on curriculum issues, while the other emphasized computer activities. The hope was that when the statements from these brainstorming sessions were combined and sorted, the computer activities would be distributed across curriculum areas throughout the map. Instead, however, participants perceived the similarity between computer activities as salient and tended to sort most computer items together. Consequently, the resulting map showed most computer items clustered together—an outcome not particularly useful if one is interested in looking at how computer activities relate to curricular areas.

To deal with this perceived semantic distinction (between computer items and curricular activities) a reanalysis of the data was undertaken. All similarity values between computer items in the N × N group similarity matrix were set equal to zero. Thus, when the map was recomputed using multidimensional scaling, computer-related items were only allowed to be related to curricular issues and not to each other. This reanalysis led to a map which was more appropriate to the issue of integrating computers into the curriculum.

Serendipitously, it seems that this analytic method may have general utility for further refining a concept map or for exploring secondary issues. It can be used whenever one is interested in seeing how subsets of items might intermix, even though they appear in different areas on the original map. For instance, if concept mapping is used for planning purposes within an organization, one might initially obtain several clusters describing separate areas of organizational activity,

Requests for reprints should be sent to Douglas Keith, 675 Sylvan Way, Redwood City, CA 94062.

with one of those clusters describing personnel issues. While this may be of interest in and of itself, it may also be desirable to explore how the individual personnel issues fall *within* other activity area clusters. This might be useful in identifying how certain personnel problems are related to other organizational activities. Even though the personnel items might be placed together as a group on the original map, because of their salience as a group, the methodology outlined here could be used to force these items to intermix with the other conceptual groups on the map.

In the following sections, the original concept mapping is described, the reanalysis is presented along with the revised map, and the implications of the reanalysis are discussed.

METHOD

Subjects and Procedure

This study was conducted with a group of 18 teachers from a school district in upstate New York, who had formed a committee to investigate computer usage within the district, and to recommend a plant for its implementation over the next several years. For the structured brainstorming session, the larger group was divided into sub-groups and asked to generate responses to two general questions, each of which was posed separately:

1. Which areas or tasks associated with teaching or with the curriculum do you find particularly boring, tedious, or repetitive; and, which curriculum areas do students typically have a hard time grasping?
2. What types of activities, in either general or specific terms, do computers perform well?

Groups were given approximately 20 minutes to brainstorm each of these questions. Open discussion and non-judgmental attitudes were encouraged during the sessions.

Items generated from questions 1 and 2 were combined for the second session. The lists were first edited to remove redundant items and to settle ambiguities; the final list consisted of 50 entities which represented those items most frequently generated during the brainstorming session. The complete list is given in Table 1. Each of the 50 entities was numbered and printed on a separate slip of paper. Each member of the group was instructed to sort all 50 entities into groups of items which he or she perceived to be similar. No constraints were imposed on the participants in regard to the dimensionality or cognitive constructs on which their various sorting strategies could be based. This sorting process was repeated three times by each participant. After each sort, participants recorded the number of separate piles of items and the identifying number of each item sorted into each pile.

The sort information was aggregated and a two-dimensional multidimensional scaling and cluster analysis were performed as described in Trochim (this volume).

Interpretation of the Map

Interpretation of the MDS plot involves a process of defining clusters of entities on the plot, meaningfully labeling those clusters, and then providing a rationale for the particular groupings and for the labels that have been given to them. Interpretations for this exercise were conducted by the author, and later by the original participants in a group session. There was a high degree of correspondence between these two interpretations; an abridged description follows of the clustering of the school-related items.

Interpretation of the plot suggests that participants perceived four conceptually distinct areas within the original set of items: (a) Teacher Management: teacher management and organizational tasks; (b) Curriculum Delivery; day to day teaching and learning activities involving the established curriculum; (c) Higher-Order Thinking: cognitively based higher-order thinking skills; and (d) Attitudes and Values; psychological factors which may or may not be addressed by the educational process, but which are crucial in some respect to that process.

Within these four regions, Clusters 6, 8, 9, and 10 appear to represent the teacher management and organizational activities: for example, creating lesson plans and materials, grading papers, reporting grades, etc., and has therefore been labeled the "Teacher Management" cluster. Cluster 10, containing items involving library skills and locating information, was apparently perceived as also being closely related to teacher management activities: it involves activities outside of the immediate classroom, and tends to involve organizational skills. Cluster 10, therefore, was also subsumed into the Teacher Management cluster.

Clusters 3 and 1, which make up the Curriculum Delivery region, are comprised almost entirely of items pertaining to specific areas of the curriculum: for example, reading skills, math concepts, social studies skills, constitution and government, etc. These are interspersed with entities associated with personal skills and study habits which are generally considered to be necessary to students' success within a formal schooling environment, such as recall of facts, general study skills, and communication skills.

Cluster 2 contains the items "following directions" and "comparison and categorization skills." In Figure 1, this cluster stands by itself under the region named Higher-Order Thinking.

TABLE 1
LIST OF STATEMENTS FROM BRAINSTORMING

1. Recall of facts
2. Application of concepts/skills; practice in applying concepts to practical situations
3. Math concepts: word problems; knowing how to set up math problems
4. Reading skills, comprehension
5. Social studies skills
6. Scientific principles
7. Attitudes and values, responsibility, citizenship
8. Organizational skills: keeping track of things, creating organizational systems
9. Study skills
10. Creating lesson plans
11. PSEN
12. Spelling: recognizing misspelled words, correcting misspelled words
13. Correcting papers
14. Reporting grades; taking attendance
15. Enrichment of basic curriculum
16. Logic and reasoning
17. Following directions; getting instructions; having to repeat instructions
18. Providing variety; encouraging variety in the curriculum, in student work
19. Review of facts
20. Computer simulations: ability to model behavior of systems, scientific principles, social situations, etc.
21. Locating or finding information or things: reference sources; parts and materials; library-related skills
22. Communication skills: listening, reading, understanding, expression
23. Cause and effect (simple and complex)
24. Charts, graphs, tables, time lines, etc. (teaching and understanding)
25. Comparison and categorization
26. Decision making skills
27. Teaching analogies; understanding analogies
28. Self image/self concept
29. Vocabulary (and meaning): commonly used vocabulary; specialized vocabulary
30. Constitution and government: Understanding social and governmental systems
31. Solid geometry
32. Fractions
33. Planning ahead: prediction, understanding option and consequences
34. Motivation: to try, to learn, to succeed
35. Creating materials; repetitive materials
36. Library related tasks: cataloguing, circulation, inventory and purchasing
37. Graphics capabilities of computers: aid to visualization, pictures, graphs, charts, animation
38. Computer drill (ability to repeatedly present simple problems and check answers
39. Potential of computers to provide positive reinforcement, praise, reassurance, suggestions, help, etc.
40. Computers (can) save paper and paperwork
41. Computer managed instruction: diagnosis of skill level; prescription/remediation; individualized instruction
42. Student-computer interaction: computer accepts input and responds; one-to-one involvement
43. Computer information storage and retrieval capabilities: can store different kinds of information; can retrieve information on demand
44. Computer telecommunications: remote communications; remote data bases and information services; remote group projects
45. Computer processing speed: saves time; makes very large tasks possible
46. Computer programmability: can perform according to instructions that are given to it
47. Computer ability to solve problems according to formulas or other instructions
48. Computer aided research, data collection and statistics
49. Word processing: spelling correction, neat printouts
50. "Hemisphericity" issues: teaching to the "right brain", visualization, intuition/creativity

In the Attitudes and Values region, Cluster 4 includes the items "attitudes and values" and "self image/self concept." Cluster 5 is somewhat more tangential to this region but includes items like "motivation" and "providing variety." Cluster 7 was included here because it emphasized attitudinal issues relating to computing: "potential of computers to provide positive reinforcement."

What has become of the computer-related items? In Figure 1, computer items are indicated by dots with circles around them while non-computer items are shown as dots alone. In the figure, all of the computer items congregated on the left side of the map, predominantly in the Teacher Management region. While this is probably a fair reflection of how the people sorted the items, it does very little to tell us about how computer activities are interrelated with curricular matters, an issue which is discussed next.

Eliminating Unwanted Semantic Groupings

One objective of this study was to represent the relationship of individual computer items to specific clusters of school and curriculum items. This objective was implied in the stated purpose of the exercise: to find ways in which computer related activities, being relatively new phenomena, could be integrated into the already existing structure of the curriculum. The salient semantic grouping among the computer items, therefore, was taken to be a confounding factor in the original map because it tended to obscure the issue of primary interest.

To correct for the semantic based clustering of the

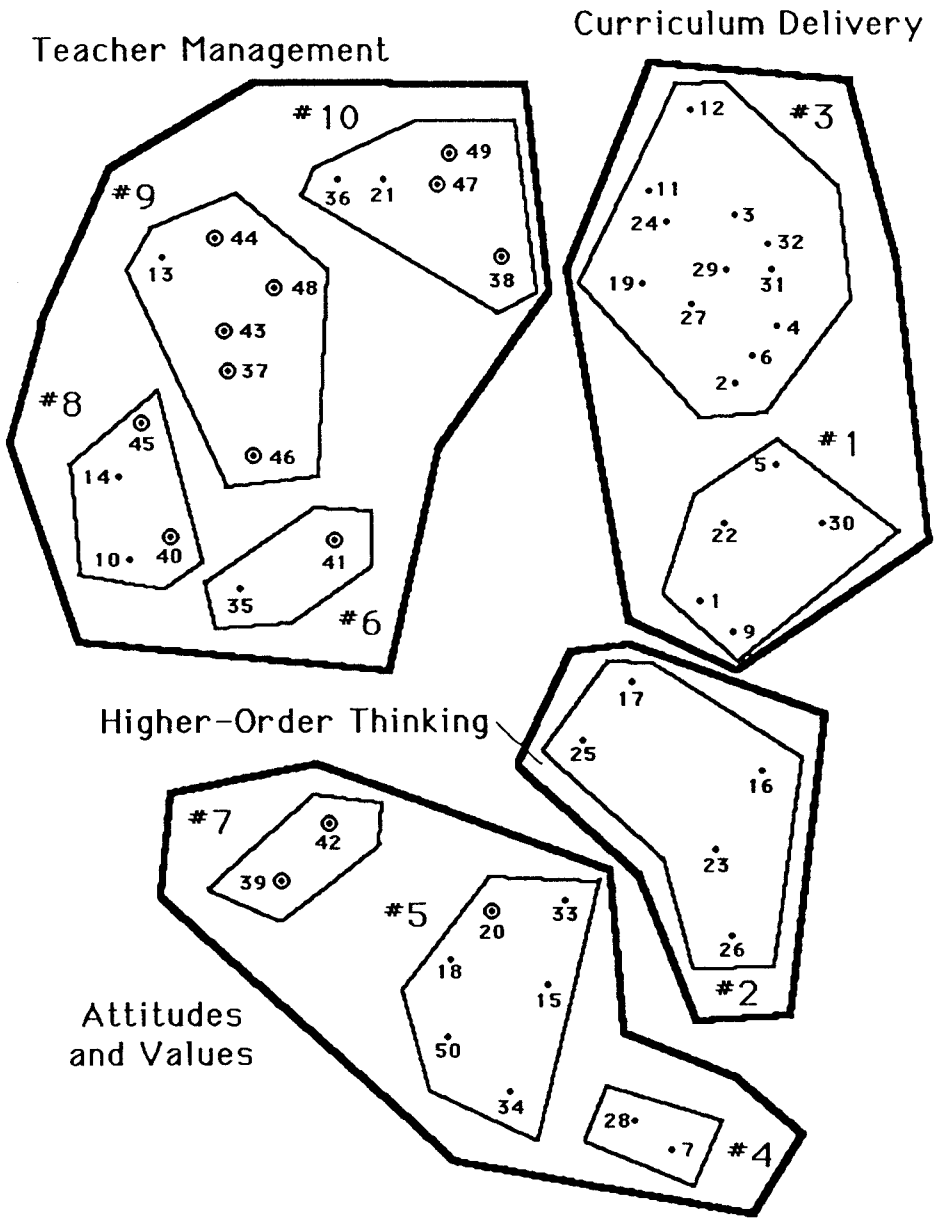


Figure 1. Initial concept map.

computer items in the original plot, an assumption was made that if the perceived relationships between all computer items were eliminated from the original similarity matrix, then the placement of those items on the map would be determined solely by their relationships to all school and curriculum items, rather than by their direct relationships to each other. It was further assumed that any clustering of computer items that might result from this manipulation of the matrix would be determined only by their similarities to various school and curriculum items. To investigate these assumptions, the cells representing all comparisons between the 14 computer items in the original input matrix were given values of zero, and the altered matrix re-submitted to the multidimensional scaling and cluster analysis procedures. The resulting plot is shown in Figure 2.

A visual inspection of Figure 2 indicates that the altered input matrix produced a restructuring of the map which more appropriately addressed the research interest. Many of the previously closely clustered computer items have been dispersed more evenly throughout the plot, and their resulting positions, in many instances, now fall into curriculum clusters in a way which appears to be more intuitively correct than in the original solution. For example, computer item 38, "drill and practice," now lies in the cluster representing spe-

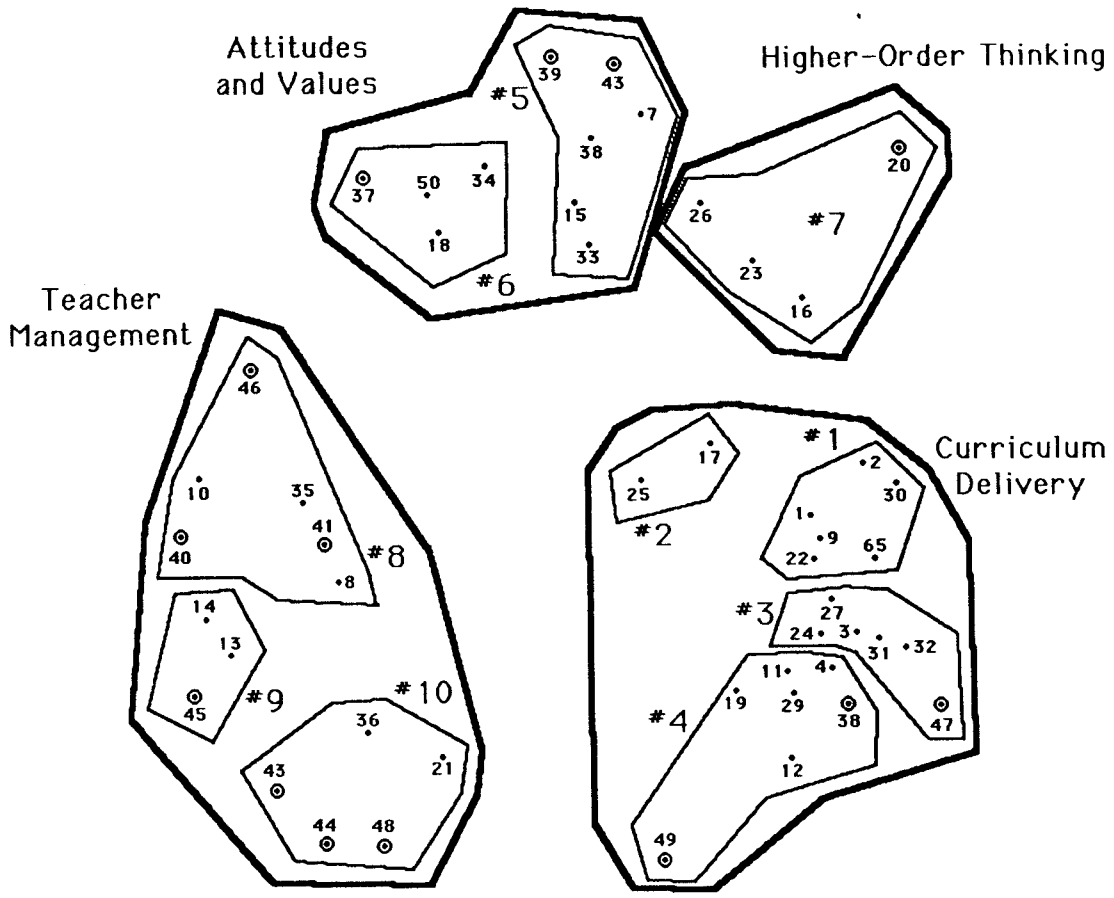


Figure 2. Revised concept map with computer-computer item similarities set to zero.

cific, factually based curriculum items. Computer item 37, "graphics capabilities," now lies in the same cluster as "hemisphericity/right-left brain issues" and "variety." Computer item 39, "computer positive reinforcement," and computer item 42, "student-computer interactions," now lie in the cluster containing "self-image" and "attitudes and values." The cluster containing items related to library activities now contains computer items 43, 44, and 48: respectively, "computer information storage and retrieval," "telecommunications," and "computer-aided research." Computer item 47, "computer problem solving capabilities," now lies in the cluster containing math curriculum items, and computer item 20, "simulations," now lies in the cluster representing higher order thinking skills.

Because the map represents the perceived similarity-distance between each entity and all other entities, zeroing out the relationships between computer items also produced some differences from the original plot in regard to the school and curriculum item clusters. These changes appear to be minimal, however, in that the integrity of most of the original clusters has been maintained, and the spatial order of the clusters remains essentially unaltered, although rotated somewhat in the two plots.

CONCLUSION

Conceptualization research of the type described in this volume is frequently concerned with representing intra-relationships among entities which may belong to distinct semantic groups. Quite often, the objective of the conceptualization exercise is to uncover hidden and implicit structure in the data set, beyond the obvious surface distinctions. Such structure is based on the nature of the free associations that subjects in the study make between entities, and for that reason it is often not desirable to impose constraints on the judgement process. On the negative side, unconstrained sorting and rating strategies often result in irrelevant semantic groupings which, at best, contribute little to the intended focus of the study.

In some cases, it may be possible to alleviate the formation of unwanted semantic groupings by attempting to ensure that the target set of conceptual entities is essentially unidimensional. If potentially unwanted groupings can be identified prior to the sorting or rating process, it may be possible to instruct participants to avoid making associations solely on the basis of those groupings. For example, it may have been possible to instruct participants in the school-computer integration study to avoid sorting items solely on the basis of their identity as school-related or computer-related entities.

The approach described here of nullifying certain inter-group associations may be primarily desirable for situations in which the multidimensional scaling output is confined to a two-dimensional solution. It is also possible that a similar effect could be achieved by specifying MDS solutions with increased dimensionality, and then examining higher orders of 2-dimensional plots; for example, dimension 2 vs. dimension 3, or dimension 2 vs. dimension 4, etc. Some of the plots may adequately describe confounding semantic groupings, while others may more accurately describe the relationships of interest. This approach remains an area for further investigation.

This paper outlines a method that was used to further explore conceptual structure in a concept mapping context by controlling for strong semantic associations which initially obscured the relationships of primary interest. While it may be best to try to eliminate such artifacts in the brainstorming step, in some cases this tactic may prove to be difficult or impossible in practice. The method outlined here might be used as a *post hoc* corrective procedure, or, perhaps more promisingly, as a tool to better refine the exploration of implicit relationships in any concept mapping process.

Evaluation and Program Planning, Vol. 12, pp. 81-86, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

VALIDITY OF MULTIDIMENSIONAL SCALING IN THE CONTEXT OF STRUCTURED CONCEPTUALIZATION

JEANNE DUMONT

Cornell University

ABSTRACT

The validity of statistical techniques for deriving accurate and trustworthy conceptual representations has received little attention in the research and evaluation literature. This paper uses hand constructed and computer maps to examine the validity of multidimensional scaling in the context of structured conceptualization. The illustration offers several strategies for examining validity. The similarity assessment reveals congruence at the level of specific results and significant differences at the level of whole map findings although the findings in total suggest the potential for overall validity of representations. Findings which have implications for planning and evaluation are discussed.

The expanding use of statistical techniques to develop conceptual representations indicates the feasibility and usefulness of such methods for deriving relational conceptualizations (see e.g., Trochim, 1985). As the role of conceptualization methods expands to include causal theory development and program assessment, it is important and timely that validity issues be examined. The lack of attention to the validity of this kind of statistical analysis to derive conceptual representations may partially be a reflection of the multiplicity of relevant yet distinct interpretations of the results of a given planning or evaluation effort.

The concept of validity here refers to the best available approximation of the truth of propositions (Cook & Campbell, 1979) and the trustworthiness of the methods and results (Guba & Lincoln, 1981). In this preliminary examination of the statistical techniques, the validity question could take several forms. Do the representations or maps reflect the ideas and experiences of participants as the participants themselves see them? Do people recognize the meaning of the maps, that is, find them sensible and interpretable? Do the maps illuminate an underlying truth or basis of a person's own theories? Do the maps reveal an underlying structure of the data?

Each of these questions involves a related but different aspect of validity. A concern in this study is whether the statistical analysis is doing what it purports to do, a concern central to the notion of construct validity. The validation process here is guided by a pattern matching approach: what is the similarity between sets of relations based upon the same objects but represented using different methods of analysis? Specifically, to what extent does a computer generated map compare with a relational map generated by hand, non-statistically?

Two ways of answering this question are attempted: (a) an assessment of the extent to which there is agreement between the spatial relationships of computed maps and hand placed maps; and (b) a record of participant impressions of the sensibility and truthfulness of the computed and hand placed maps. In this research, as in many planning and evaluation efforts, the participating individuals are guiding the propositions. Their ideas about the content—the relationships expressed in the construction of the hand placed maps—are viewed as theoretical relationships. The validity task involves determining the extent to which the computed maps, operationalizations of observed relationships, agree with participants' theoretical relationships, operationalized as the hand constructed maps. The participant reports serve a secondary illuminating function in this paper.

Requests for reprints should be sent to Jeanne Dumont, Department of Human Service Studies, MVR Hall, Cornell University, Ithaca, NY 14853.

METHOD

A method was designed to assess the statistical techniques used in a structured conceptualization about deinstitutionalization issues. Five persons who had participated in the larger study (Dumont, in progress) were selected to be involved in the verification process and construct a relational map by hand, in effect simulating multidimensional scaling, and to discuss the meaning and value of the computed representations or maps.

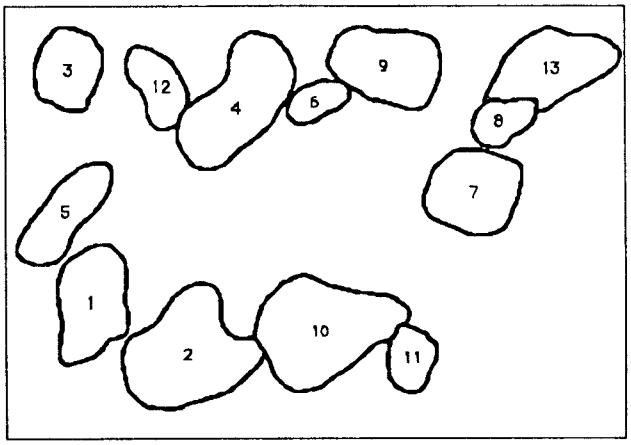
The larger study included the conducting of ten semi-structured interviews to investigate factors contributing to living in the community and reinstitutionalization from the perspective of persons psychiatrically institutionalized two or more times. The interview transcripts were coded using a documentary coding method (Wrightson, 1976) to identify cause and effect statement pairs, a condition for conducting cognitive mapping analysis. Each of the ten persons then participated in a unstructured sorting procedure; they were asked to sort their own statements several times "in a way that makes sense to you."

The sort data were combined in a similarity matrix and subjected to multidimensional scaling (Davison, 1983; Kruskal & Wish, 1978) and then hierarchical cluster analysis (Everitt, 1980). The results of the multidimensional scaling (MDS) analysis were used to make a number map representing all coded statements relationally as points in two dimensional space. The cluster analysis was superimposed on the MDS results. Decisions about number of clusters were based on statistical criteria and ease of interpretation. The results from the MDS and cluster analyses were used to make a concept map where lines or boundaries are drawn around individual statement points to represent similarly perceived statements as clusters.

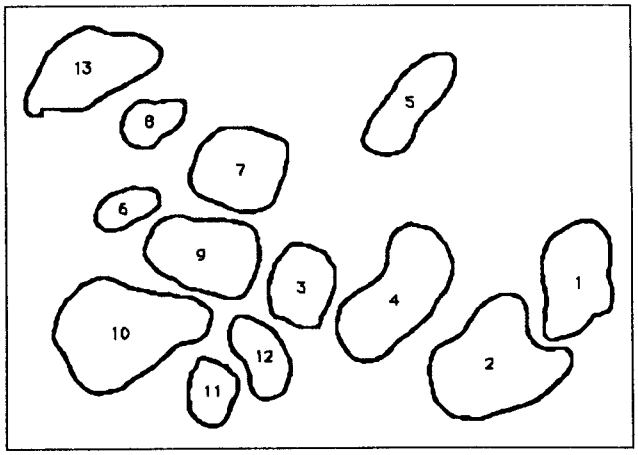
The verification checks were conducted systematically as part of a post-interview with each participant. First participants were handed a set of their interview statements, grouped into clusters, and asked to look them over. Second they were asked to place the clusters on a 32" by 40" piece of cardboard to create a relational map, i.e., place clusters viewed as most similar closest in space and clusters viewed as dissimilar farthest apart. Third, they were asked to name each cluster. Naming regional areas was suggested. Lastly, they were asked to clarify (e.g., names of clusters, etc.). After completing the hand mapping portion of the verification check, participants were shown two computed representations: (a) the number map which shows how interview statements were placed by multidimensional scaling; and (b) the cluster or concept map which shows how individual entities or statements were grouped into clusters. A brief explanation of the method of analysis accompanied the presentation of each map. He or she was given a listing of statements as they were grouped into clusters by the cluster analysis to help interpret the maps and name the clusters. The person was asked to discuss the sensibility and truthfulness or correctness of the maps and to comment on the usefulness of the maps and conceptualization process.

Upon completion of the five verification checks, the hand placed maps and computed maps were placed on the same scales. Cluster midpoints were estimated to use in computing Euclidean distances between cluster pairs. Figure 1 is an example of one participant's rescaled hand placed and computed concept maps. It is included here to illustrate patterns of agreement detectable by visual comparison.

For example, the proximity of some clusters on the hand placed and computed map is similar in a regional sense: clusters 1, 2, 3, 4, and 12 are located in the same



Theoretical (Hand Placed) Map



Obtained (Computed) Map

Figure 1. Example of a hand-placed and computed map for person 1.

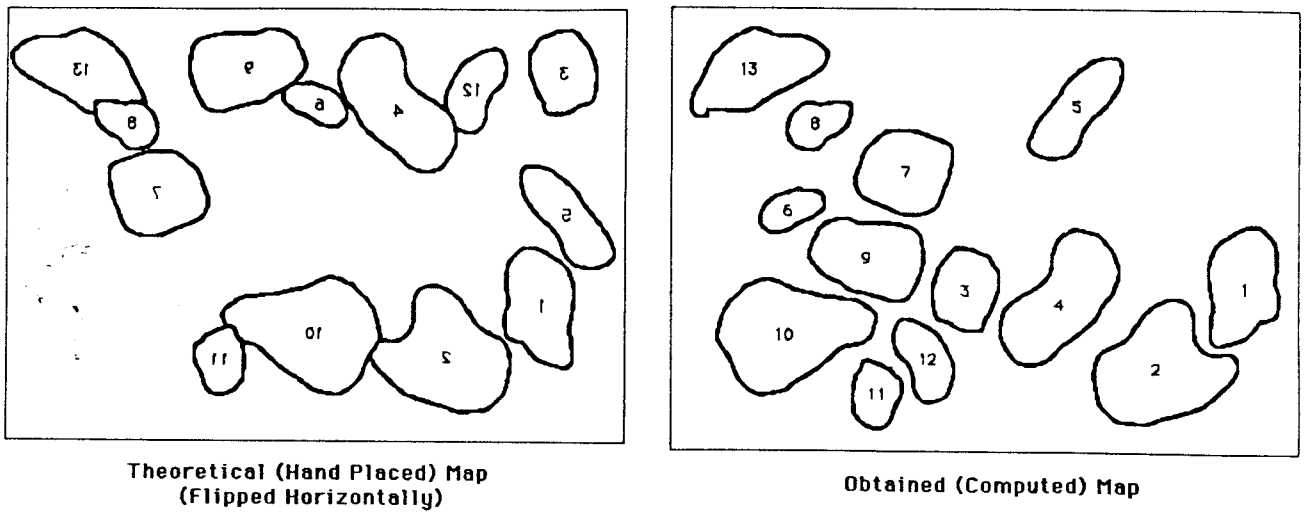


Figure 2. Example of a "flipped" hand-placed and computed map for person 1.

half of each representation. This degree of agreement is further illustrated in Figure 2 in which the hand placed map has been flipped horizontally. The hand placed map has been flipped horizontally to more dramatically demonstrate the configural similarity between the two maps. Spatial structure is unaffected by flipping or rotating a relational map.

Measuring Agreement between the Computed and Hand Placed Maps

The configural similarity comparison of the computed and hand placed concept maps is estimated using Pearson's Product Moment Correlation of Euclidian distances between points on the map (Davison, 1983; Kruskal & Wish, 1968). This correlation coefficient is a measure of congruence; in this context it is an indicator of how similar the distances are between clusters at different levels of analysis for the two maps. The points in this computation are based on the calculated midpoint coordinate values of each cluster and this method is illustrated in Figure 3. The set of distance pairs for each representation or map provides the basis for analyzing the degree of agreement between two maps. Higher coefficients at any level of analysis indicate more agreement between maps.

FINDINGS

The results of the configural similarity assessment for the whole map and regional areas are presented in Table 1. The observed correlation for the whole map configurations of person 1 is significant at $p < .001$ and indicates a fair to good measure of agreement. No other correlation at this level of analysis was as high or significant. As pointed out, Figure 2 illustrates pictorially this relatively high degree of agreement.

Initially, the whole map results for persons 2, 3, 4, and 5 might seem discouraging, the low correlations providing little evidence for a correspondence between maps. When the restrictiveness of the assumptions for computing this measure is considered (e.g., every pair of cluster distances, $N(N-1)/2$, must be compared), it may be that a low correlation results more from this restrictive assumption than from low validity of multidimensional scaling.

This paper includes a regional and cluster level correlation analysis to focus on the similarity of specific cluster patterns within the overall configurations. An intuitive observation of the maps of person 4 (Figure 4) illustrates this point at the regional level.

The overall map correlation of .24 indicates a less than fair measure of agreement between these maps. However, when certain configural patterns are examined, some similarity between the maps is evident. For example, the relative positions of clusters 11 and 12 and of 1, 3, and 8 are similar on both maps. When the five pairs of maps are examined on the regional area level (focusing on pairs of regions), nearly half of the correlations are significant (of the 13 applicable findings, 6 are significant at $p < .02$). The less stringent assumptions at this level, that is, the fewer patterns to be matched, helps to explain the higher correlation values and subsequent significant findings. However, a small sample size would usually lead to lower correlations. At this level of analysis, the less stringent assumptions alone do not account for the higher correlation values.

There are several curious correlation coefficient findings when examined at the inter-regional level (focusing

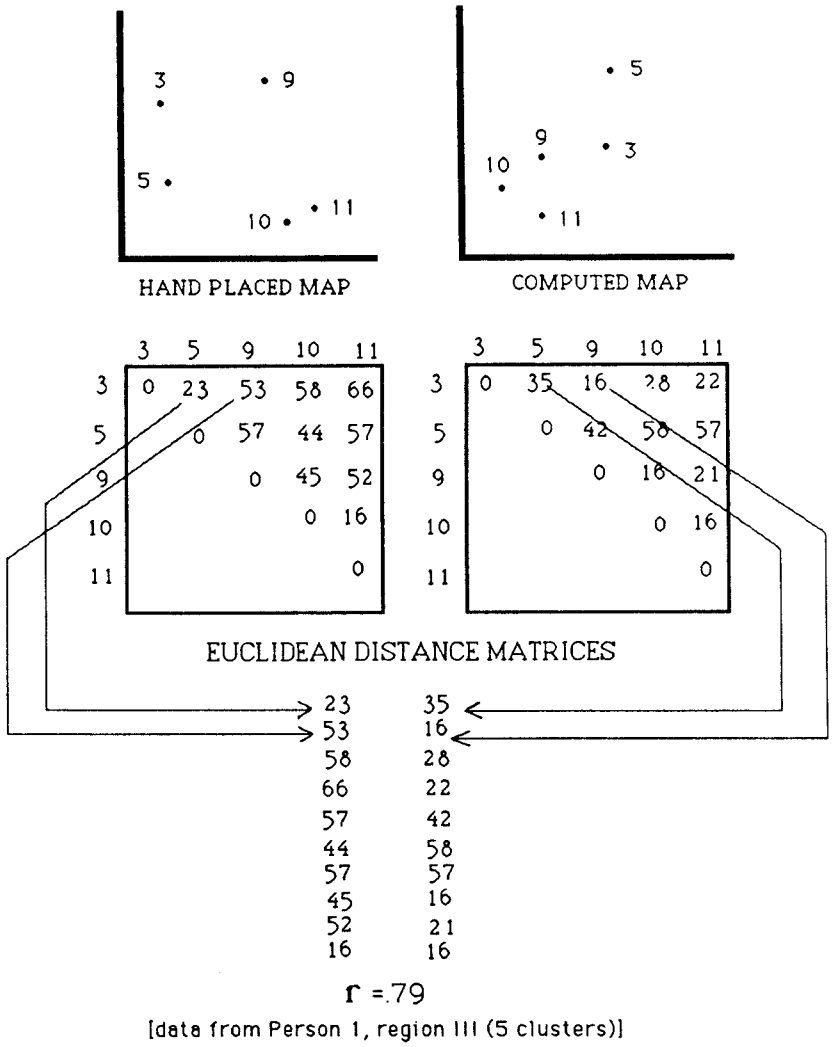
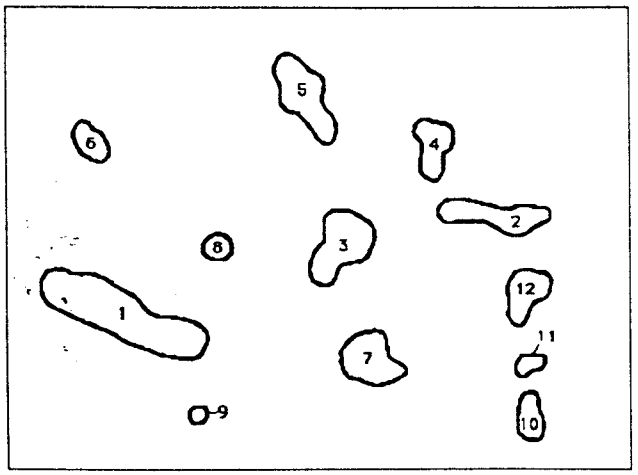


Figure 3. Method of calculating the configural similarity between two maps.

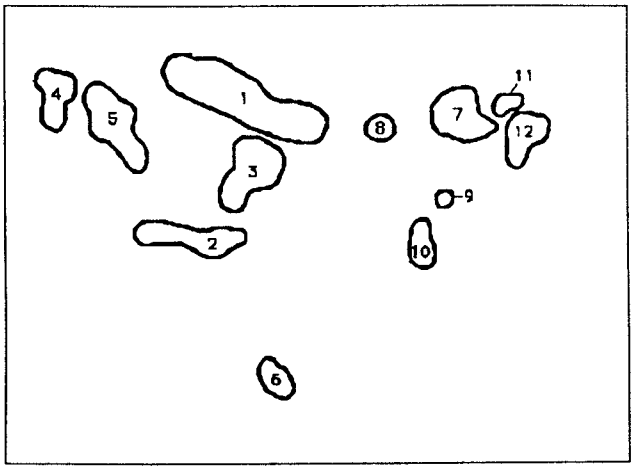
TABLE 1
CONFIGURAL SIMILARITY BETWEEN THEORETICAL AND OBSERVED DISTANCE PATTERNS USING PEARSON'S COEFFICIENT

	Person 1	Person 2	Person 3	Person 4	Person 5
Whole Map	.50*** (91)	.02 (91)	.15 (78)	−.24 (78)	−.10 (105)
Inter-Regions	.81 (3)	.81 (3)	—	−.89 (3)	−.34 (3)
Regional Areas					
I & II	.44* (8)	.28 (8)	.15 (12)	.31 (8)	.02 (10)
I & III	.35* (10)	.15 (9)	NA —	.50** (8)	.01 (10)
II & III	.70*** (8)	.04 (9)	NA —	.47* (8)	.72*** (8)
Within Regions					
I	.22 (5)	−.58 (4)	.60* (6)	−.35 (4)	−.27 (6)
II	.60 (3)	.23 (4)	−.34 (6)	.19 (4)	.63 (4)
III	.79** (5)	.44 (5)	NA —	.18 (4)	.28 (4)

n's are shown in parentheses.
*statistically significant at the .02 level.
**statistically significant at the .01 level.
***statistically significant at the .001 level.



Theoretical (Hand Placed) Map



Obtained (Computed) Map

Figure 4. Example of a hand-placed and computed map for person 4.

on the centroids of regions). This level of analysis can assist in determining whether a low correlation is a measurement or theoretical problem and what additional measures should be employed to account for the findings. For example, person 1's correlations were statistically significant at every level except the inter-regional level. The observed correlation value between the three centroids in this case may be insignificant because the critical r value for significance at one degree of freedom is exceptionally high. Divergent correlations at the regional or cluster levels can specify where one ought to look in verifying the validity of the method.

Table 2 shows the correlation coefficients for individual cluster relationships, based on the Euclidean distance between each cluster and every other cluster. It shows Pearson's r between each set of distances for the hand placed and computed map when focusing on one cluster at a time. In other words, the degree of association between the hand placed and computed map has been calculated for the set of row values taken from the entire matrix of Euclidean distance pairs. If, for example, there are 14 clusters as with person 5, each correlation coefficient in Table 2 is the degree of association among the 13 pairs of distance relationships when focusing on one cluster at a time. The correlation values for each cluster of the computed and hand placed comparisons are all significant at the .02 level or higher. The less restrictive assumptions cannot alone account for the significance of the findings. The consistently high correlation values suggest multidimensional scaling does a good job of representing inter-cluster relations when focusing on one cluster at a time as focal pairs.

CONCLUSIONS

Several preliminary conclusions can be drawn concerning the validity of multidimensional scaling in the context of structured conceptualization. First, certain patterns in each participant's hand placed concept map were in agreement with the computed concept map. These patterns of configural similarity cut across the level of relationships examined and reasonably approximate some relationships at each level.

Second, the validity of the whole map analysis was impressively supported by the findings of one case (person 1). This participant articulated the importance of formulating a system of beliefs and emphasized the value of the computed maps and conceptualization process itself. Obviously an individual's system of beliefs can promote evidence for whole map validity in this pattern matching approach. The credibility and usefulness of the process and outcome to the participant may also (in)directly support the validity of the whole map findings. The findings suggest the potential for overall validity of representations yet uncertainty about the degree to which the statistical techniques can, with any consistency, lead to pictorial representations on the whole map level that adequately reflect, perhaps reveal, existing viewpoints and perspectives.

Third, notwithstanding that many representations may well reflect the thinking of an individual or group at a particular point in time, some representations are less inaccurate and more trustworthy than others. It's important to discern the extent to which inadequate agreement in configural similarity between hand placed and computed concept maps reflects limitations of the method as compared with limitations of the theory for examining the relationship among concepts. This analysis suggests that dimensionality constraints may, for example, impose a spatial limitation in measurement that may be contributing to the disparate results. Or, it may be that multidimensional scaling better represents certain similarity relationships or cognitive styles (e.g.,

TABLE 2
CONFIGURAL SIMILARITY OF CLUSTER RELATIONSHIPS USING PEARSON'S CORRELATION COEFFICIENT

	Person 1	Person 2	Person 3	Person 4	Person 5
Cluster 1	.964***	.799**	.914***	.893***	.648*
Cluster 2	.948***	.896***	.935***	.762**	.689**
Cluster 3	.901***	.777**	.928***	.909***	.835***
Cluster 4	.913***	.865***	.783**	.899***	.832***
Cluster 5	.899***	.844***	.870***	.943***	.682**
Cluster 6	.906***	.778**	.728**	.904***	.908***
Cluster 7	.907***	.848***	.696*	.896***	.772**
Cluster 8	.946***	.793**	.886***	.812**	.928***
Cluster 9	.914***	.862***	.834***	.864***	.936***
Cluster 10	.840***	.792**	.900***	.847***	.879***
Cluster 11	.865***	.899***	.808**	.845***	.868***
Cluster 12	.870***	.890***	.812	.790**	.851***
Cluster 13	.963***	.865***	NA	NA	.845***
Cluster 14	NA	NA	NA	NA	.780***
Average	.910	.839	.841	.864	.818

*statistically significant at the .02 level.
**statistically significant at the .01 level.
***statistically significant at the .001 level.

multidimensional scaling represents a kind of similarity that is more complementary with relational thinkers than linear thinkers but may serve a greater heuristic function for predominately linear thinkers). An analysis of validity which relies on a statistical assessment, as this effort does, is limited in examining the interrelationship between the perspective of the analyses and participant thinking dynamics.

Further empirical study is needed to assess the degree of validity using multidimensional scaling for conceptualizing viewpoints or perspectives, including program theory. The potential unaccounted bias of representing similarity as closeness in 2-dimensional space is a potential limitation which raises questions about the validity of relying upon distance information alone to construct program effect patterns. If the focus of a conceptualization is on the process itself (e.g., internal organizational development), then such a limitation may be relatively insignificant. An assessment of the relative validity of using multidimensional scaling alone and with other methods can lead to further understanding of how limitations may eventually be minimized. Such assessments could benefit from a mixed method approach.

REFERENCES

COOK, T.D., & CAMPBELL, D.T. (1979). *Quasi-experimentation: design & analysis issues for field settings*. Chicago, IL: Rand McNally College Publishing Company.

DAVISON, M.L. (1983). *Multidimensional scaling*. New York: John Wiley.

EVERITT, B. (1980). *Cluster analysis*. New York: Halstead Press.

GUBA, E.G., & LINCOLN, Y.S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.

KRUSKAL, J.B., & WISH, M. (1978). *Multidimensional Scaling*. Beverly Hills, CA: Sage.

TROCHIM, W.M.K., & DUMONT, J.M. (1985, October). *Combining the use of multidimensional scaling, cluster analysis and cognitive mapping for conceptualizing evaluations*. Paper presented at Evaluation '85: A Joint Meeting of the Canadian Evaluation Society, Evaluation Network, and Evaluation Research Society, Toronto, Ontario, Canada.

TROCHIM, W.M.K., & LINTON, J. (1986). Conceptualization for planning and evaluation. *Evaluation and Program Planning, 9*, 289-308.

TROCHIM, W.M.K. (1985). Pattern matching, construct validity, and conceptualization in program evaluation. *Evaluation Review, 9*(5), 575-604.

WRIGHTSON, (1976). Documentary coding method. In R. Axelrod, *Structure of decision: The cognitive maps of political elites*. Princeton, NJ: Princeton University Press.

Evaluation and Program Planning, Vol. 12, pp. 87-110, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

CONCEPT MAPPING

Soft Science or Hard Art?

WILLIAM M.K. TROCHIM

Cornell University

ABSTRACT

Is concept mapping "science" or "art"? Can we legitimately claim that concept maps represent reality, or are they primarily suggestive devices which might stimulate new ways to look at our experiences? Here, the scientific side of concept mapping is viewed as "soft science" and the artistic one as "hard art" to imply that the process has some qualities of both, but probably does not fall exclusively within either's domain. In the spirit of hard art, a "gallery" of final concept maps from twenty projects is presented, partly to illustrate more examples of the process when used in a variety of subject areas and for different purposes, and partly for their aesthetic value alone. In the spirit of soft science, two major issues are considered. First, the evidence for the validity and reliability of concept mapping is introduced, along with some suggestions for further research which might be undertaken to examine those characteristics. Second, the role of concept mapping is discussed, with special emphasis on its use in a pattern matching framework.

One of the lingering questions regarding concept mapping is: should it be considered primarily a "scientific" process or an "artistic" one? Without getting into the voluminous literature on the distinctions and commonalities between art and science, we might at the least assume that science strives for objectivity—for an agreed upon consensus about reality—whereas art is more subjective in nature—striving to reach and challenge our perspective on reality in a unique, personal manner. Perhaps even posing the issue dichotomously obscures some of the commonalities between art and science. After all, like art, science requires creativity and insight especially in theory development stages. And like science, art requires a mastery of skills, methods, tools, and technologies.

This paper does not provide an answer to the question. As with most issues which are dichotomously framed, it may very well be that the truth of the matter lies somewhere in the middle. Concept mapping may result in both a representation of reality and an interesting suggestive device. In anticipation of such a non-conclusion the title of this paper has been tempered somewhat, describing the scientific side of concept mapping as "soft science" and the artistic one as "hard art" to imply that the process has some qualities of both, but probably does not fall exclusively within either artistic or scientific domains. This paper will examine the concept mapping process from each perspective—viewing it as an artistic procedure which yields interpretable, suggestive conceptual pictures and as a scientific one based upon sound evidence regarding its validity, reliability, and theory-enhancing value.

The question of whether concept mapping is primarily hard art or soft science is an important one on several counts. First, it is essential that those who use the process not misrepresent it to clients. It is easy to believe that the maps which result have a quality of finality to them—they imply objectivity and scientific truth; they seem to represent some group conceptual reality. If this is not true—and our clients believe that is—then we are deceiving them and practicing scientism of a most insidious type. Second, concept mapping would have different uses, depending on how we choose to answer the question. If it is a credible scien-

Requests for reprints should be sent to William M.K. Trochim, Department of Human Service Studies, MVR Hall, Cornell University, Ithaca, NY 14853.

tific procedure, we might use it as the basis for developing theory, as a way to develop theoretical patterns for pattern matching, or as the basis for generalizability of inferences (Trochim, in press). If concept maps are not justifiable as accurate representations of some underlying reality, they might alternatively be considered primarily as suggestive devices, useful for their stimulative or creative value. In this latter case, the focus would be more aesthetic in nature—how do we generate interesting maps which suggest new ways of looking at familiar problems? Finally, the way in which we evaluate the effects of concept mapping would differ depending on whether it is hard art or soft science. From a scientific viewpoint, we would want to know the degree to which concept maps are reliable, valid and parsimonious in their representation of theory or data. From an artistic point of view we would be more concerned about the aesthetic judgment of the participants—the degree to which the map is pleasing or interpretable—and we would look at the effects of the process on stimulating new ideas, or facilitating subsequent planning or evaluation efforts.

The next section is in the spirit of artistry and aesthetics and presents a brief gallery of concept maps from a variety of different projects. These maps provide more examples of the concept mapping process and will illustrate a variety of techniques for graphing concept maps. The section after that presents some of the scientific issues which have been raised and suggests some of the research which needs to be included on our agenda for further investigations of the concept mapping process.

CONCEPT MAPPING AS HARD ART: A GALLERY OF PROJECTS

As with all art, beauty is in the eye of the beholder. Because concept maps are the product of a fairly complex set of group interactions, it is reasonable to expect that even for people within the participating group, appreciation and understanding of a concept map will differ. If so, then it is almost certainly the case that someone who did not participate will have more difficulty in understanding what a concept map means to the participants. Every mapping process involves a unique group of individuals who come to the process with their own experiences, jargon, and motivations. The outside viewers of a map are not privy to this and, instead, apprehend the picture with their own perspectives and biases. It is worth recognizing, then, that there is the effect of the concept map upon the participant group and organization(s) and its effect upon some outside viewer of it, and that these two may not be in agreement.

Here, we review briefly the final concept maps from twenty different projects (only one of these has already been seen in the introductory chapter of this volume). These projects are included here for a number of reasons. They illustrate a wider variety of subject areas for the application of the concept mapping process. They show different graphing techniques which can be used to display concept maps. They depict different purposes and uses of concept mapping. But perhaps most of all, you may find that some of the maps are simply "pretty" and interesting to look at. Unfortunately, only final maps can be presented here due to space limitations. More detailed descriptions of any project may be requested from the author.

For several projects, the anonymity of the organization is preserved (usually at their request). The twenty projects and their major characteristics are described in Table 1. They range over a five year period beginning in the summer of 1983 and with the most recent ones conducted in the spring of 1988. They range in size from projects with as few as 4 participants to as many as 75; as few as 11 statements or as many as 137. The participants include organizational staff, board members and client groups. These projects represent an evolution in our understanding of concept mapping from some of the cruder early attempts to more sophisticated recent ones.

The twenty projects are classified by subject areas in Table 2. Most of the projects are related in some way to education, but there are also other fields well represented. Some projects are pertinent to multiple subject areas. For instance, project 10, the Community School of Music and Arts, is classified in four subject areas—education, educational administration, children and youth, and the arts—even though its primary emphasis was on the improvement of administration of the school. The projects are classified by purpose in Table 3. The vast majority of projects were conducted for planning and/or management purposes.

Although all of these projects generally followed the procedure described in the introductory chapter of this volume, there were slight variations over time. For instance, early projects did not incorporate rating data into the maps. Also, the quality of map production generally improved over time. Each project will be described briefly and its final concept map presented.

Multicultural Awareness Camp

In the summer of 1983 a group of youth workers constructed a camp program for local high school children. But this day camp had a unique twist to it—it was designed to help the children become more aware of different groups and cultures, and to raise issues of

TABLE 1
CHARACTERISTICS OF PROJECTS

	Date	Number of Persons	Types of Participants	Number of Statements
1. Multicultural Awareness	Summer 83	12*	core staff	74
2. Division of Campus Life	Winter 83	45*	staff – all levels	137
3. University Health Services	Spring 84	50–75	all staff	100
4. DCL Subcommittee	Fall 84	11	dept. representatives	11
5. Mental Health Association	Fall 84	8–12	board members	80
6. Teaching Measurement	Fall 85	20*	graduate students	50
7. Cooperative Extension	Fall 85	---	staff	75
8. Student Life (R.A.s)	Fall 85	12	undergraduate students	129
9. Student Life (dorm residents)	Spring 86	10	undergraduate students	46
10. Community School of Music and Arts	Summer 86	10	board members	54
11. Alumni Affairs	Summer 86	10–15	staff	72
12. Student Life (Student Assembly)	Summer 86	20*	undergraduate students	87
13. Employment	Summer 86	8–10	agency representatives	90
14. Personnel Management	Fall 86	4	all staff	74
15. Counseling Services	Fall 86	10–15	staff and board	80
16. Senior Citizens	Spring 87	10–15	staff and board	75
17. Elderly	Winter 87	10–15	agency representatives	95
18. Arts Council	Winter 87	10–15	staff and board	63
19. Planned Parenthood	Spring 88	30–35	staff and board	89
20. Music and Arts in Daycare	Spring 88	10–15	daycare providers	61

*approximate numbers participating.

TABLE 2
CLASSIFICATION OF PROJECTS BY SUBJECT

	Education	Educational Admin.	Children & Youth	Mental Health	Elderly	Health	The Arts
1. Multicultural Awareness	X		X				
2. Division of Campus Life		X					
3. University Health Services						X	
4. DCL Subcommittee		X					
5. Mental Health Association				X			
6. Teaching Measurement	X						
7. Cooperative Extension							
8. Student Life (R.A.s)	X	X					
9. Student Life (dorm residents)	X	X					
10. Community School of Music and Arts	X	X	X				X
11. Alumni Affairs	X	X					
12. Student Life (Student Assembly)	X	X					
13. Employment							
14. Personnel Management							
15. Counseling Services				X	X	X	
16. Senior Citizens					X		
17. Elderly					X		
18. Arts Council	X	X					X
19. Planned Parenthood				X		X	
20. Music and Arts in Daycare	X	X	X				X

class, race, gender and sexual orientation for individual and group thought, discussion and action. The day camp was an intensive $4\frac{1}{2}$ day program with ongoing follow-up in the school and community.

The purpose of this project was to develop a map of the staff's goals for the program so that the staff might be better able to understand what the students might get from it and how they themselves as staff might benefit. To generate the statements for concept mapping four core staff members responded to two questions:

TABLE 3
CLASSIFICATION OF PROJECTS BY PURPOSE

	Planning	Evaluation	Survey Design	Curriculum Development	Theory Building	Management
1. Multicultural Awareness	X	X				X
2. Division of Campus Life	X					
3. University Health Services	X					X
4. DCL Subcommittee	X					X
5. Mental Health Association		X		X		
6. Teaching Measurement					X	
7. Cooperative Extension	X			X	X	
8. Student Life (R.A.s)	X				X	
9. Student Life (dorm residents)					X	
10. Community School of Music and Arts	X			X		X
11. Alumni Affairs	X			X		
12. Student Life (Student Assembly)	X					
13. Employment	X					X
14. Personnel Management	X					X
15. Counseling Services	X					X
16. Senior Citizens	X					X
17. Elderly	X					X
18. Arts Council	X					X
19. Planned Parenthood	X					X
20. Music and Arts in Daycare	X	X	X	X		

1. What are your personal hopes for the multicultural program in terms of what it will accomplish?
2. How do you want the adolescents' lives to be different during the camp and afterwards?

The final cluster map for the 74 generated statements is shown in Figure 1. The map seems divisible into three separate regions of clusters, one each for participant effects, school and community effects, and effects of the program on the staff. The participant effects are, not surprisingly, the most detailed, with a total of four clusters grouped in the upper left of the map, one each for the effects on: individual students (e.g., reach a clearer sense of self; feel safe to explore difficult or confusing topics); communication skills (e.g., talk to each other in more respectful, less hurtful ways); ongoing action (e.g., continue having these kinds of dialogues with each other and other adults); and, action

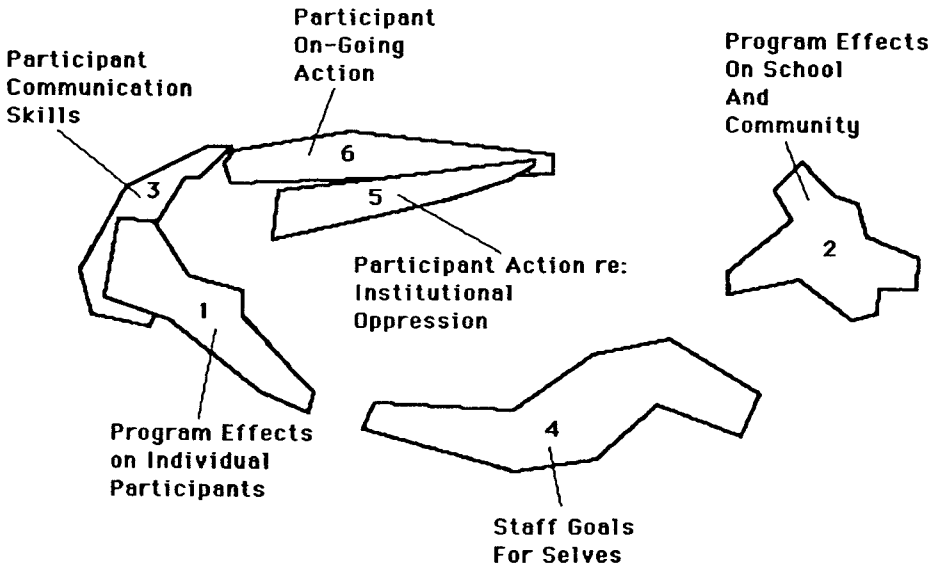


Figure 1. Final concept map for the Multicultural Awareness Project.

regarding institutional oppression (e.g., learn effective ways of challenging institutional oppression with support from each other). Staff effects are shown in the cluster at the bottom of the map (e.g., want to know what kinds of approaches and processes work best in enabling young people and adults to explore issues of racism, sexism, etc.). The effects of the program on the school and community (e.g., teachers will examine the kinds of information which gets transmitted to kids) are shown in the cluster on the right of the map.

The map gave the staff a visual image of what they hoped to achieve and helped assure that all staff had some common sense of what the program was about. The map was also used as the basis of a qualitative interview assessment of the program conducted subsequently.

Division of Campus Life

The Division of Campus Life (DCL) is an administrative unit at Cornell University responsible for delivering a great variety of services (e.g., student residences, transportation, safety, dining, counseling, health, etc.) to the University community. It is comprised of eleven different departments which vary according to size, organizational structure, and type of function performed. The goal of this project was to produce a map which could be used as an organizing device for the long-range planning effort of the DCL. This project and the broader planning effort have been described in greater detail in Trochim and Linton (1986) and Gurowitz, Trochim, and Kramer (1988).

This was a very large project compared with the others reported here. At any given stage, there were approximately 45 people involved representing the eleven departments. The focus for the brainstorming was the mission statement of the DCL. The statement logically divided into three major phrases and one brainstorming session was held for each. Because of the number of people, 876 statements were originally brainstormed. This was clearly too many and so a subcommittee of four participants was appointed to examine the set of statements for redundancies and they reduced them to a final set of 137. The final map is shown in Figure 2. The map was divided into four general regions. On the left side, and considerably distant from the other three regions is the one labelled "human development and values." Most of the items which fell

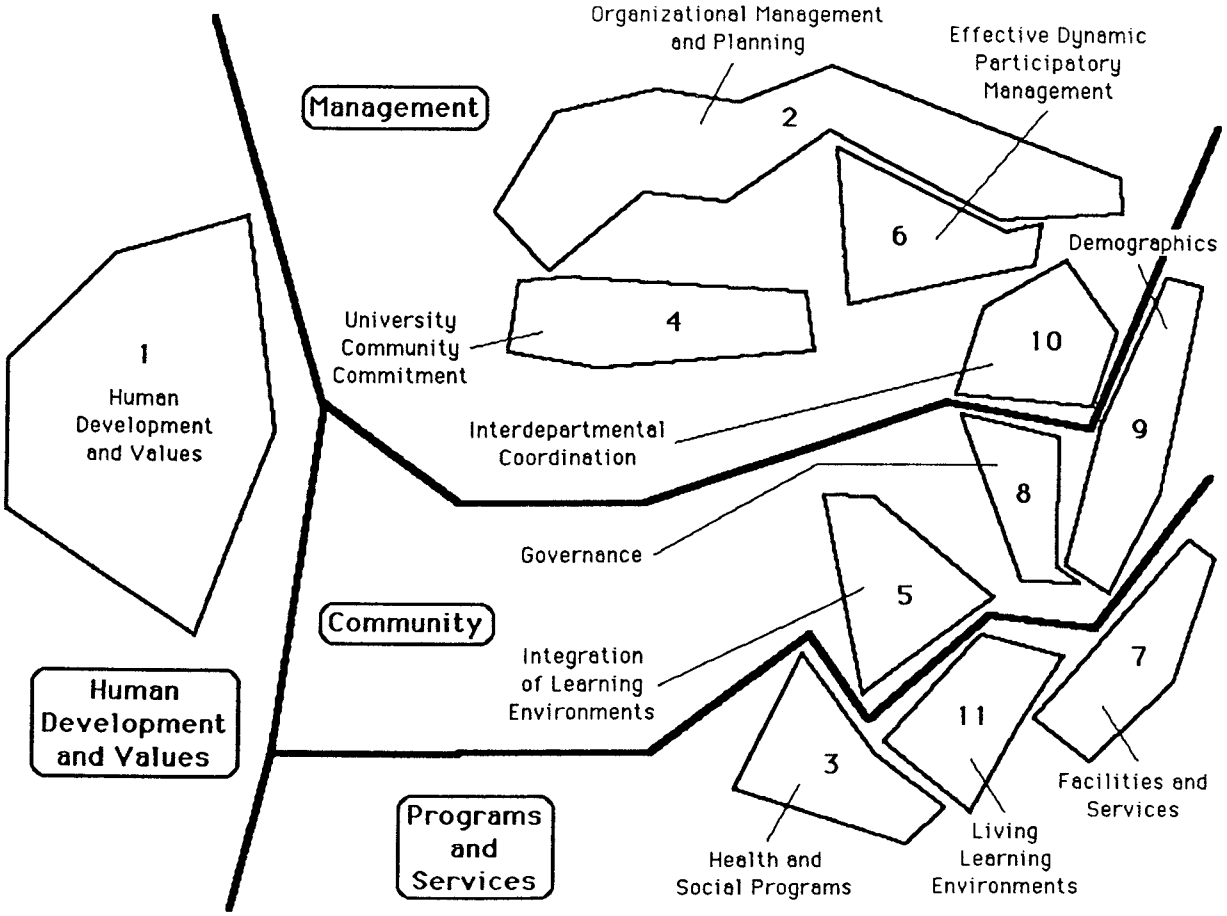


Figure 2. Final concept map for the Division of Campus Life Project.

into this category were short general statements. The three regions on the right (management, community, programs and services) contained statements which tended to be longer and more concrete in nature. One reason for this left-to-right split might be because of the three-part focus (based on the mission statement) which was used. One part of the mission statement seemed to call for more general value statements while other parts implied more concrete actions.

The map for this project formed the basis of subsequent long-range planning as described in Trochim and Linton (1986) and Gurowitz et al. (1988). In addition, the next two projects described below were direct offshoots of this original project and were also connected with DCL planning.

University Health Services

This project was conducted to provide a framework for long-range planning for the University Health Services (UHS) which is a department in the Division of Campus Life described above. At any given stage there were between 50 and 75 participants, with virtually everyone in the department participating in at least one stage. For the brainstorming, participants were asked to generate statements which describe what the UHS "should be or should do." Because of the number of participants, three separate brainstorming sessions were held with about 25 persons at each. A total of 315 statements were generated and 100 were randomly selected from these to produce the final set. The map which resulted is presented in Figure 3. The map is divided into four regions and twelve clusters. In this map, directions appear to be directly interpretable. Moving from the top to the bottom implies moving from more managerial or administrative issues to ones that are more educational or service related. Movement from left to right denotes a change from external, service and client-related issues to the more internal concerns of the staff. To utilize the map, the participants were divided into small groups which were responsible for generating specific recommendations for action (action statements) for different regions and clusters within regions. As a result, 145 action statements were generated and each was addressed by the planning committee in subsequent meetings. This project is described more fully in Trochim and Linton (1986).

One especially interesting interaction occurred in connection with the naming of Cluster 11: Meetings and Scheduling. This cluster contained a number of statements which reflected staff dissatisfaction with the

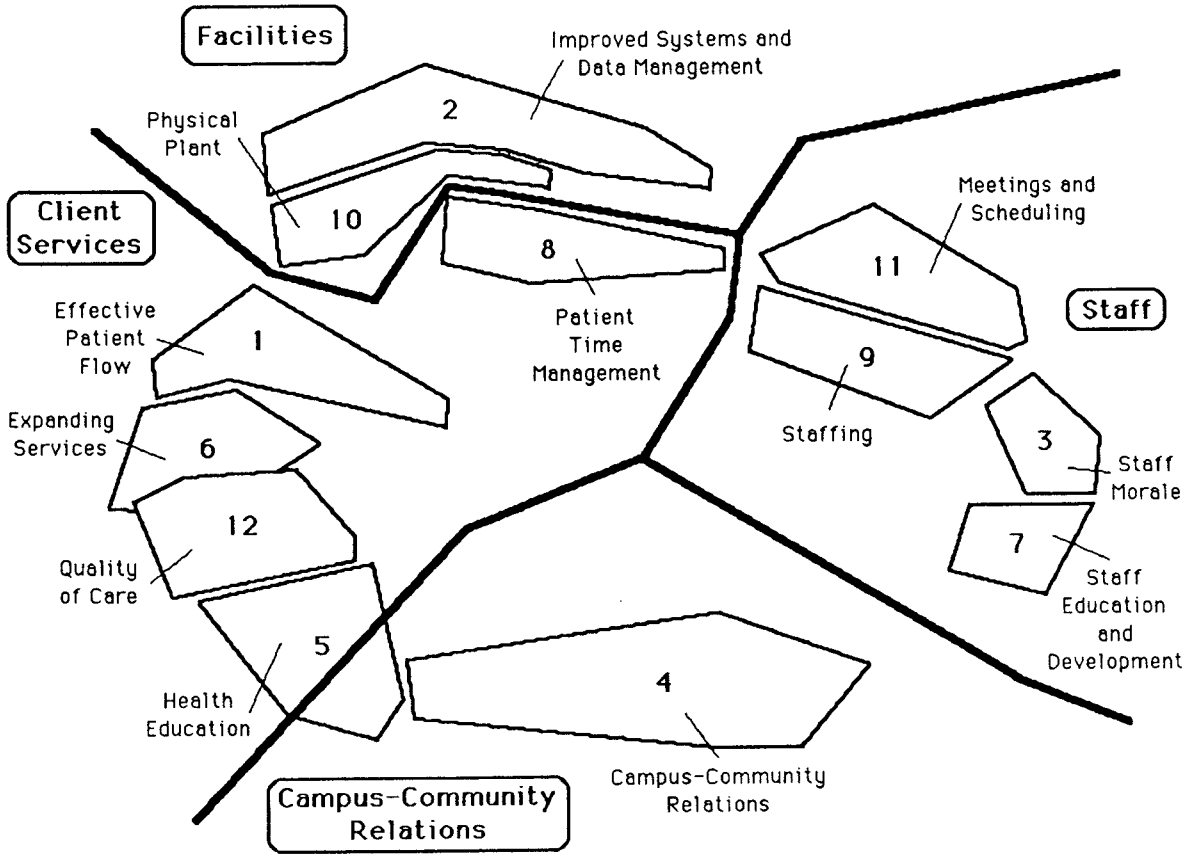


Figure 3. Final concept map for the University Health Services Project.

current arrangement of staggered lunch times (thereby preventing some staff from going to lunch together) and frequent required "sack lunch" meetings (thereby preventing them from leaving the building). For whatever reason, the staff had not previously felt comfortable in stating their displeasure with these arrangements, perhaps because they did not feel powerful enough to affect the policies. During the naming of the clusters—a session in which everyone in the organization (including the director and manager) participated—one of the participants kiddingly suggested that the cluster be named "Eat—not meet!" There was an immediate and vigorous round of applause to this tongue-in-cheek (but very heartfelt) suggestion. Afterwards the director confided that he had had no idea that there was a problem—from his perspective, the lunch times had been staggered to provide continuous coverage and the "sack lunches" had been added to provide more inservice for staff—both goals which he assumed the staff supported. The cluster naming session enabled the staff to make their feelings about these policies felt in a very direct way and, at the same time, preserved their anonymity and mitigated their fears about confronting their bosses.

DCL Subcommittee

Following the major concept mapping process for the DCL as described earlier, a subcommittee was formed whose responsibility it was to write the long-range plan for the Division. This subcommittee consisted of eleven members, one representing each department. One of the first issues which they needed to address was the question of whether all of the eleven departments in the Division actually made sense there or whether some should be moved under another branch of the University administration. Obviously, long-range planning could not proceed until some agreement was reached about what departments were in the Division. In addition, it seemed to subcommittee members that some departments had more in common than others and that they could be involved in more active collaboration, planning, staff interaction and exchange, and so on.

To examine these issues, we decided to do a concept mapping of the eleven departments to see how the subcommittee members would group them. Instead of using brainstorming to generate statements, the names of the eleven departments were used. Instead of sorting the statements, all pairs of departments were rated on a scale from 1 (least similar) to 100 (most similar) and these ratings were standardized and averaged across subcommittee members. The map which resulted is shown in the top part of Figure 4. The map shows some interesting distinctions. The departments on the left are by far the largest in terms of budgets, number of FTEs, and so on. They also tend to be devoted to providing basic services such as transportation, health, safety, food (dining), and general goods (campus store). Departments on the right of the map emphasize services and counseling. The departments in the lower right corner in general tend to deal primarily with students whereas other departments tend to serve the broader university community. The major reorganization question centered around whether the three departments on the far right should be moved to a different administrative part of the university where they would be with other departments which provided services and counseling directly to students. In the subcommittee discussions, it became apparent that most members were content with the current structure—there were two advocates for reorganization, both of whom represented student service departments on the right of the map. The two advocates were outspoken, articulate proponents of their position—most others in the group did not reveal a strong position in the face of this advocacy. After many lengthy meetings discussing reorganization, we decided to use the data from the concept map in Figure 4 to construct a "person" map to see whether it might help move the issue along. Again, we used the rating data, but this time we scaled people (using the INDSCAL model for multidimensional scaling). The map which resulted is shown on the bottom of Figure 4. The identities of the subcommittee members were not indicated on the map—only arbitrary ID numbers known only to the facilitator were used. What was particularly remarkable was that everyone in the group immediately identified the two persons in the lower right corner correctly as the two strong advocates for reorganization. There was a laugh of recognition, very little discussion, and *the reorganization issue was never seriously raised again*! While it is purely conjecture, it seems reasonable that the person map made it clear that there was a strong majority opposed to reorganizing and the two vocal advocates simply dropped the issue given that they were clearly outnumbered.

The Mental Health Association

This project was undertaken to provide a framework for designing a training program for volunteers who were to work on a one-to-one basis with deinstitutionalized mental patients. A new staff person had been hired for this purpose by the Advocacy Committee of the Board. The dilemma which faced this new staff member was that it was not clear what the Advocacy Committee wanted the volunteers trained to do. If the staff member went ahead and designed something that seemed sensible, it might not adequately represent the original thinking of the Board Members. Therefore, the Advocacy Committee was encouraged to use concept mapping as a way to represent their thinking for the new staff member. The map which resulted is shown in Figure 5. The training program was constructed so that each cluster was represented in the sessions. Individual

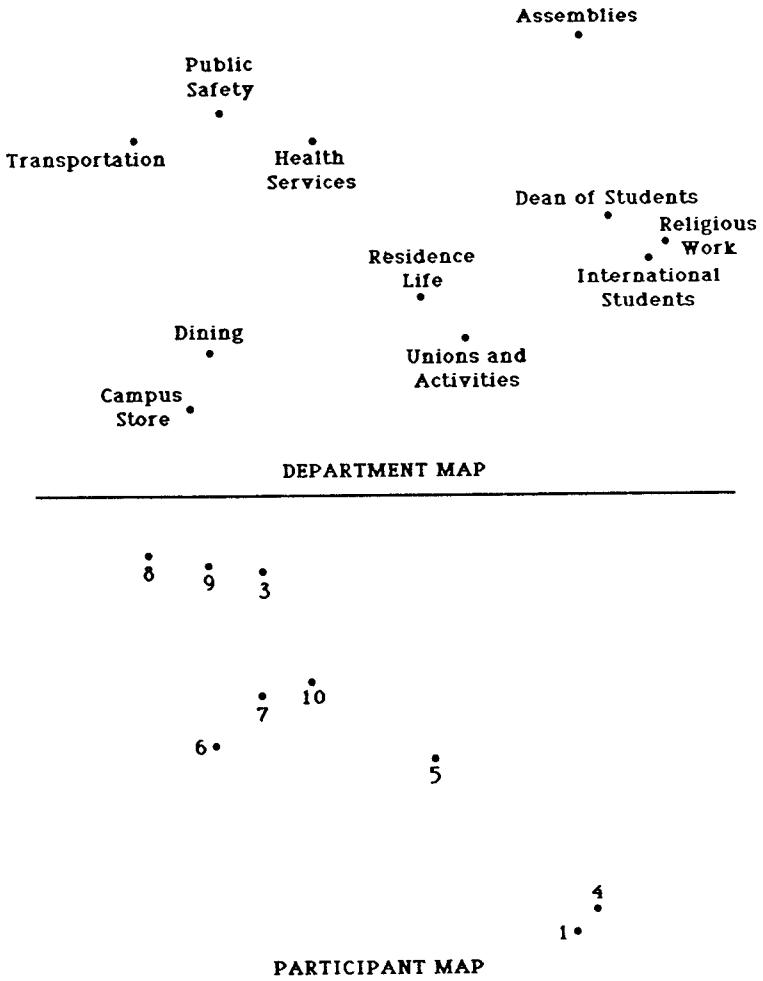


Figure 4. Final concept map for the DCL Subcommittee Project.

statements within clusters were used to suggest specific skills which needed to be taught. For instance, the cluster Supportive Communication Skills had the statements "listen attentively," and "be comfortable with hanging out together without talking." The cluster Assisting Toward Getting Resources had the statement "help clients find housing." Each of these statements was covered in some way in the training sessions. Although it wasn't done here, the same map structure could be used to evaluate the implementation and effectiveness of the program. In this project, consequently, the new staff member was able to solicit specific, concrete guidance from the relevant Board members regarding the nature of the program as they saw it—certainly better than trying to second-guess what they might have wanted.

Teaching Measurement

This project was accomplished in two class sessions as part of a graduate level course in measurement. It was conducted early in the semester in order to determine what the group perceived as the major issues in measurement and the interrelationships between these issues. Students were prompted simply to generate statements which describe what they thought about "measurement." The map which resulted is shown in Figure 6. The students identified six clusters for the 50 statements. What is especially interesting is that they perceived a counter-clockwise pattern across clusters which described the measurement process from beginning to end. Measurement begins on the far left of the map with the cluster Theory/Conceptualization (e.g., theoretical framework), then to Practical Considerations (e.g., how time consuming, budgetary factors), onto Tools of Measurement and Data Collection (e.g., surveys, questionnaires, etc.), to "Scaling and Testing" (e.g., ranking or ordering things), to "Quality of Measures/Analysis" (e.g., reliability, validity, precision), and finally onto Presentation (e.g., summary, recommendations, publishing). This process showed the students that they already (collectively) knew a considerable amount about measurement, and at the same time it

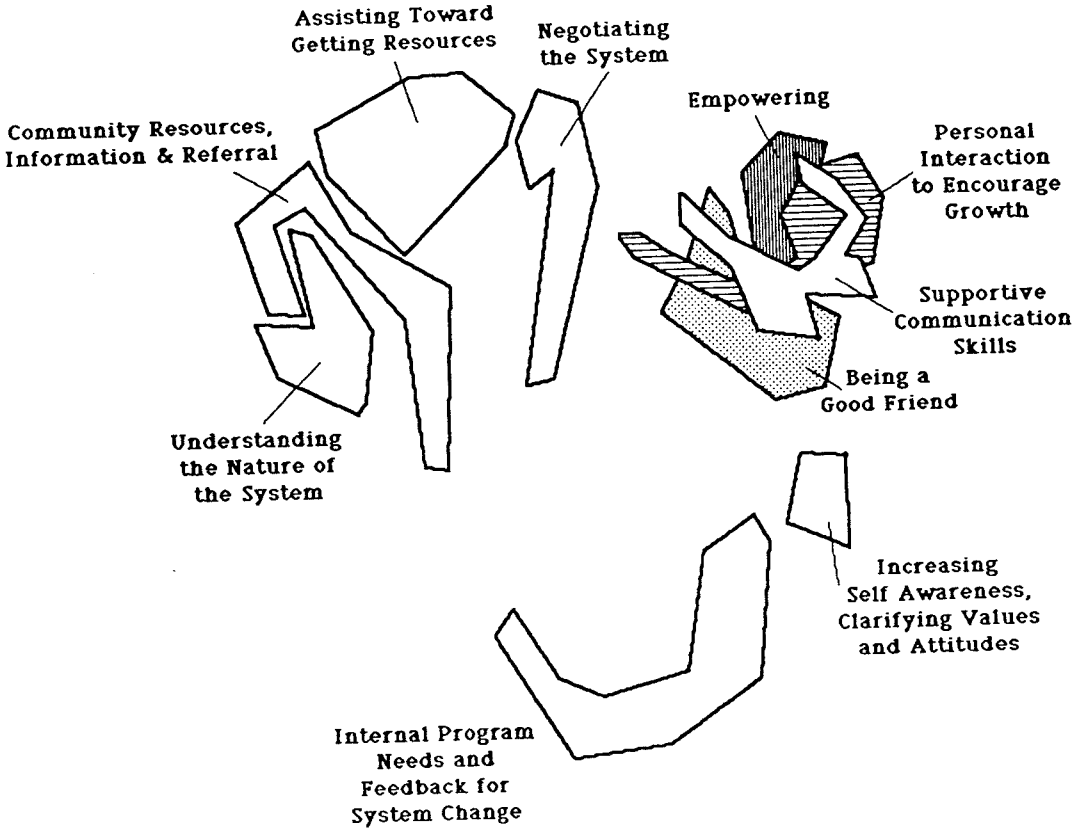


Figure 5. Final concept map for the Mental Health Association Project.

introduced them to the ideas of multidimensional scaling and cluster analysis—topics which were covered later in the course.

Cooperative Extension

Cooperative Extension has a rich history in community issues programming (CIP) which has become known as a specialty often separate from the more traditional subject areas of agriculture, home economics and youth development (4-H). Many of the community issues selected for educational programs have been in land use, agriculture districts, economic development and local government. While these are important areas for local policy making, topics more germane to home economics or human ecology have not been addressed by most CIP specialists despite the fact that Cooperative Extension home economics agents have undertaken many programs that include a "community issues" component (e.g., daycare, housing for the elderly, water quality, and consumer rights).

A steering committee was appointed to strengthen the awareness of community issues within home economics subject areas and to provide a strategy for increasing the skills and resources of faculty and agents who have a focus on educating families to incorporate more policy education into this subject area programming. Concept mapping was used to facilitate the planning process of the steering committee; to provide a benchmark for future evaluation; and, to represent the perceptions of county agents, faculty at the state university with at least 50% extension time, and administrators. Participants were randomly selected from subsets of Cooperative Extension personnel whose main focus is in home economics or human ecology programs. They were asked to brainstorm ideas which "describe your view of what a policy education program for individuals and families in Cooperative Extension should be." A total of 273 items was generated, with 75 randomly selected for use in concept mapping. Items were mailed to participants for sorting. The steering committee was responsible for interpreting the final map shown in Figure 7.

The map is divided into nine clusters: Citizens and Policy ("citizens learn how to analyze concerns" and "...impact on political process"); Governmental Processes ("state government legislative process," "local government decision-making..."); Research and Policy Analysis ("What are the likely effects on decision outcomes from different forms of citizen participation?"); Local Program Delivery ("Explain how program can help individuals and families"); Cooperative Extension Programming ("Packaged' ideas for

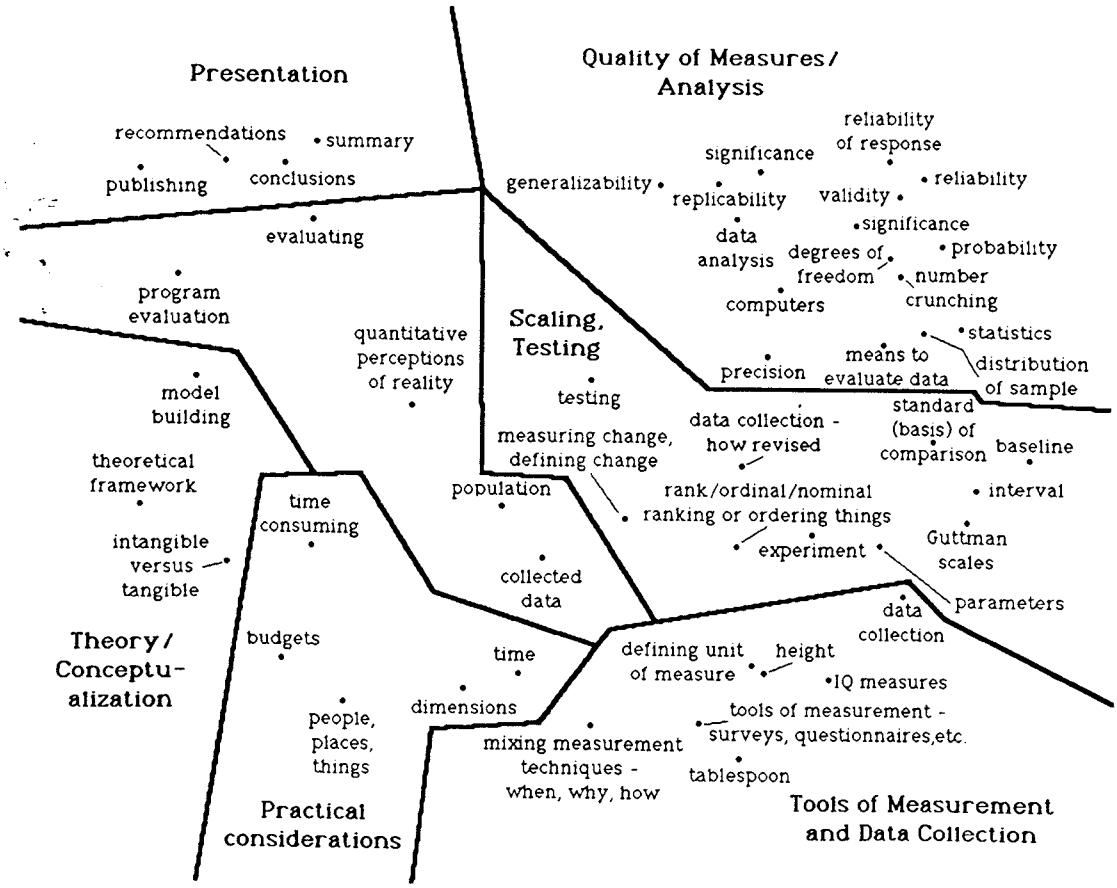


Figure 6. Final concept map for the Teaching Measurement Project.

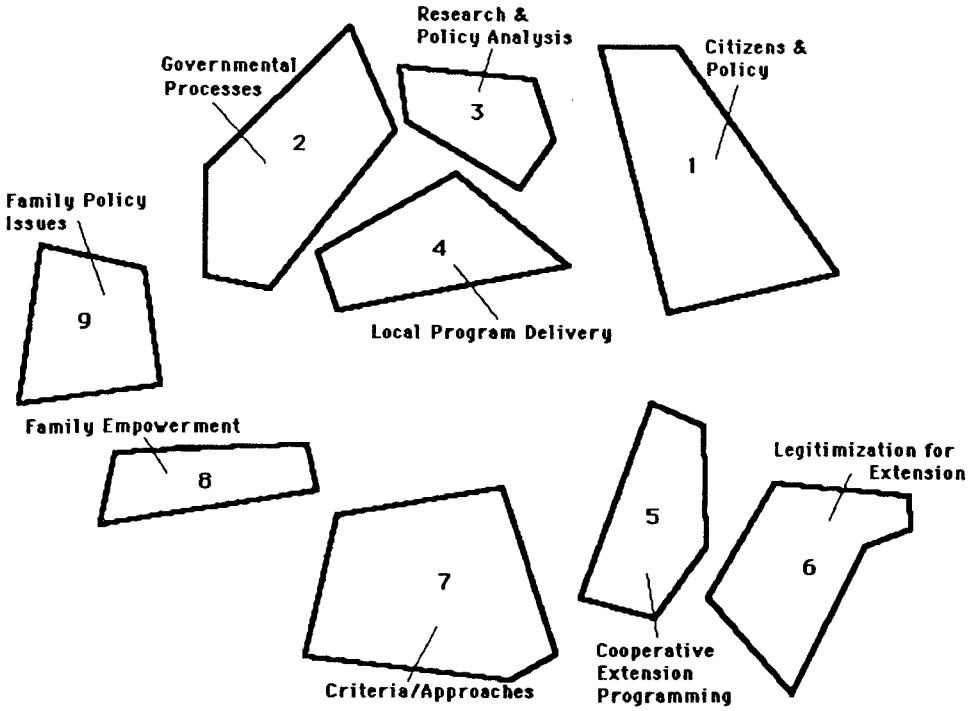


Figure 7. Final concept map for the Cooperative Extension Project.

agents"); Legitimization for Extension ("aim to develop public awareness," "success stories from other counties"); Criteria/Approaches ("program should aim at pre-school, youth, adult and aging populations in the context of the family"); Family Empowerment ("helps families understand tax problems," "parent forum: address problems of today and how handled"); and, Family Policy Issues ("support of child care," "environmental issues: water policy, safety"). The map is useful for specific planning efforts. For instance, one committee member suggested that the cluster categories be used as guides for developing materials for Cooperative Extension staff to have for inservice education opportunities and reference. The specific statements within clusters would provide concrete examples of the topics which materials could address.

Student Life (R.A.s)

This concept mapping project was conducted to explore the issues which college students perceive as important. Twelve Residence Advisors (R.A.s)—students who live in the dormitories and have some supervisory and peer counseling responsibilities—were engaged as participants. The R.A.s were asked to generate statements which "represent what you perceive to be important in the lives of undergraduate students." A total of 129 statements was generated. It was hoped that the concept map would provide a framework which would help the R.A.s to plan better and more comprehensive programs for their students. The final map is shown in Figure 8.

The map is divided into three general regions (Academic Issues, Life Issues, and Self and Others) and 15

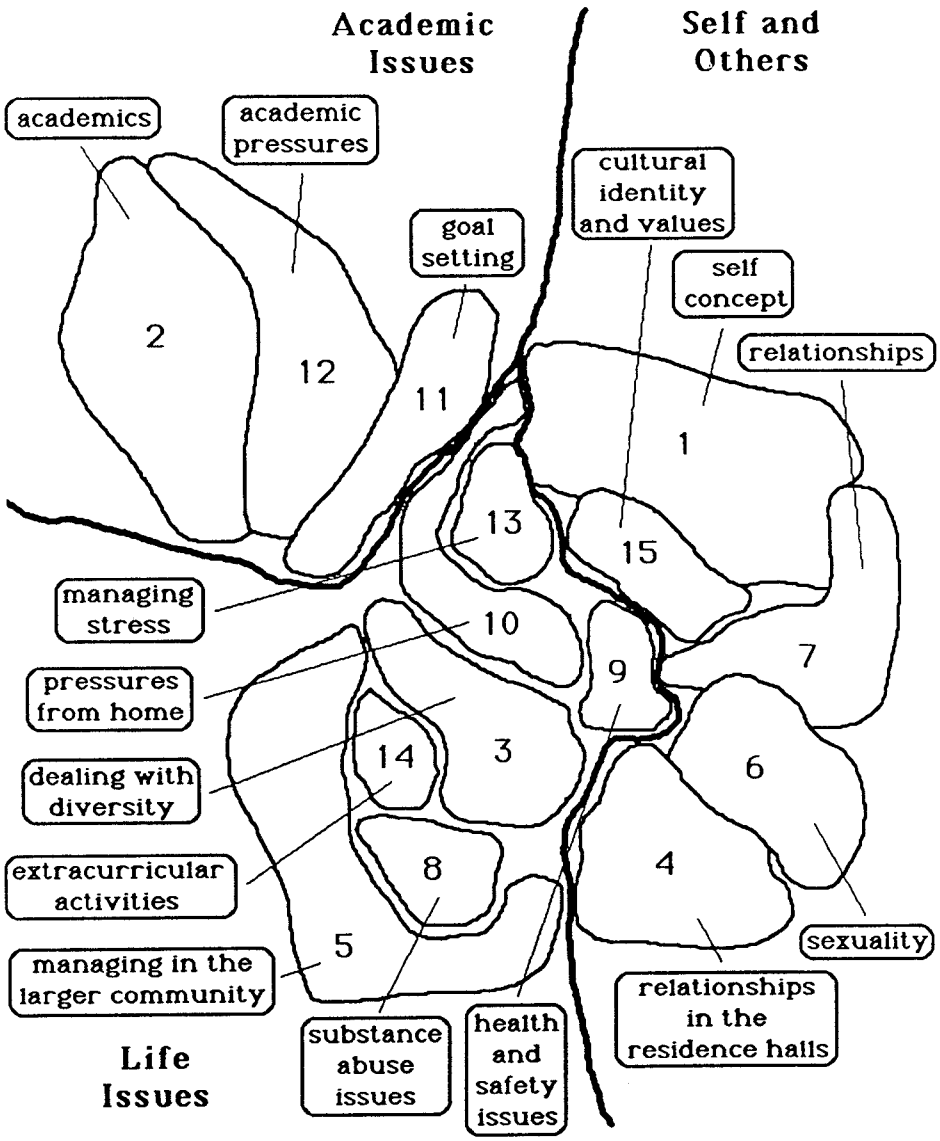


Figure 8. Final concept map for the Student Life (R.A.s) Project.

clusters. There are a number of uses for the map. R.A.s could examine currently planned programs and mark all areas on the map which are addressed in some way. This would show, at a glance, which topics are emphasized and which, if any, are neglected. The specific statements within clusters might provide concrete suggestions for what types of programs might be useful. The map may also suggest interesting relationships which could be emphasized in programs. For instance, the clusters Goal Setting, Managing Stress, and Pressures From Home are close together on the map. This suggests that the topics are strongly related and it might be valuable to construct programs which attempt to get at this. Finally, the map itself might be an interesting device for stimulating discussion among students regarding the issues which are important for them. In a sense, the map could be used as the basis for a discussion program.

Student Life (Dorm Residents)

Like the previous study, this one was undertaken to explore what issues students see as important in their lives. Ten students were asked to brainstorm statements which "describe the issues in your lives as students." This study was also designed to examine what kind of map might result from a relatively short process, and consequently, the students were stopped when they had generated only 46 statements. The map which resulted is shown in Figure 9.

What is perhaps most striking is the degree to which this map resembles the structure of the map in the previous student life study shown in Figure 8. The three regions were very similar: Identity and Social Issues (Figure 9) is similar to Self and Others (Figure 8); School (Figure 9) is similar to Academic Issues (Figure 8); and Outside University (Figure 9) is similar to Life Issues (Figure 8). Even at the cluster level there seems to be a great deal of similarity. For instance, in both figures clusters related to stress turned up near the center of the map. In fact, if the two maps are rotated (while preserving the interrelational structure) it is possible to actually overlay the two figures and achieve a fairly good correspondence. This occurs in spite of the fact that the studies were conducted with two different

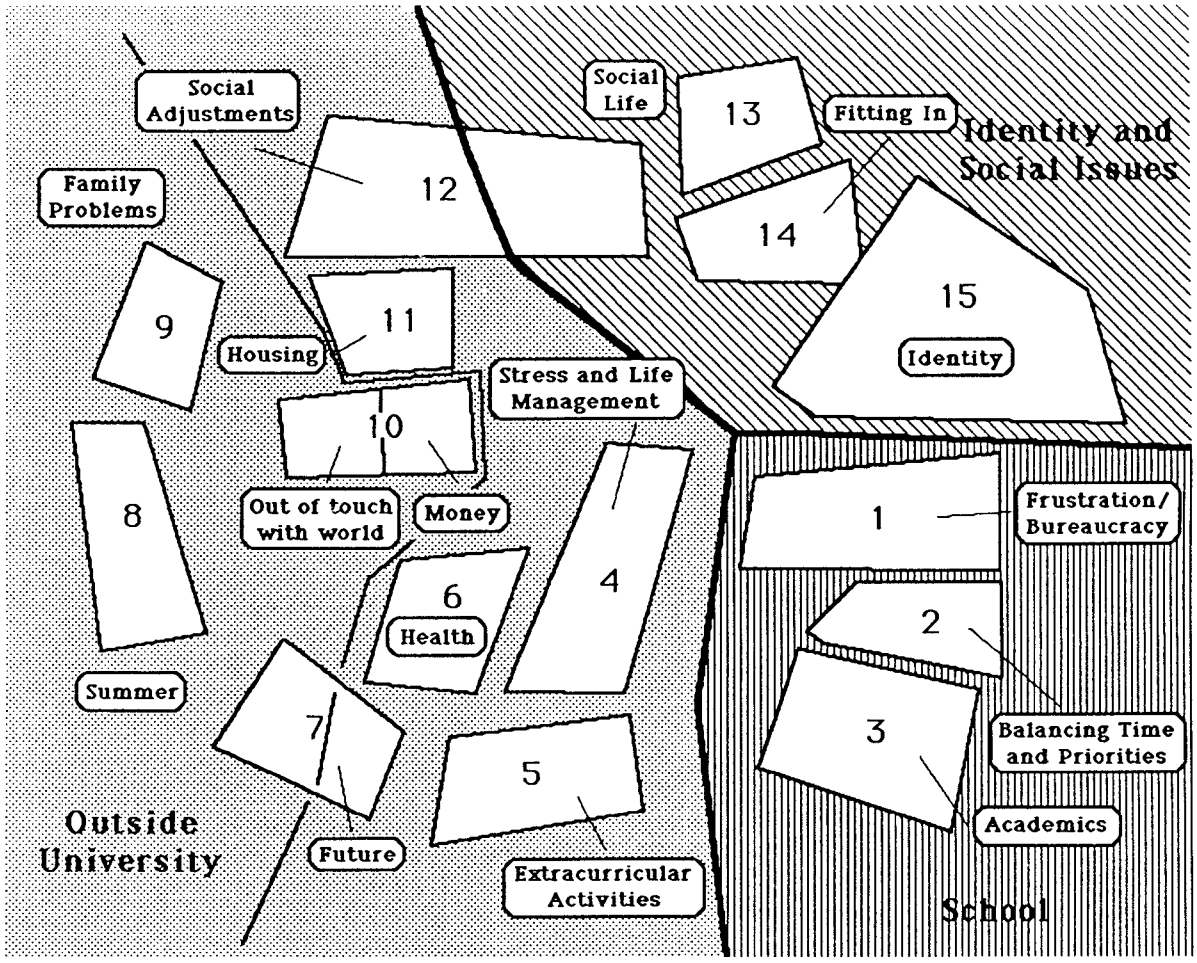


Figure 9. Final concept map for the Student Life (dorm residents) Project.

ps of students—the older R.A.s and the younger lents—and that they involved completely indepen- processes (students in one were not aware of the r) from brainstorming through interpretation. This spondence in concept maps may suggest that there me generalizable, consistent conceptual similar- hich would be useful for building educational the- a topic which will be discussed in greater detail r.

mmunity School of Music and Arts

Community School of Music and Arts (CSMA) is source for instruction in music, art and dance, and center for cultural activity in the community which rves. The goal of this concept mapping process was nvolve the Board of Trustees in the development of amework for CSMA programs and services. Ten mbers of the Board of Trustees participated in two sions. To generate the statements they used the focus mpt: "Knowing what you know about the Commu- School, identify items which will help insure its continued growth." Fifty-four statements were generated and the final map which resulted is shown in Figure 10. There were four major regions—Organization; Finances; Program; and, Benefits, Equipment and Physical Plant—and ten clusters. The map was used to plan for program development and fund raising. In addition, one of the major uses of the map was to encourage greater unity and cohesiveness between the trustees and faculty as they jointly participated in additional planning.

Alumni Affairs

The alumni Regional Directors supervise offices throughout the country which maintain contact with and provide services to graduates of a major university. This project was designed to help them conceptualize the major issues which their jobs and offices address. Between 10 and 15 regional directors generated 72 statements which describe "activities and services which your office provides or might provide." The final map, shown in Figure 11, is divided into five major regions

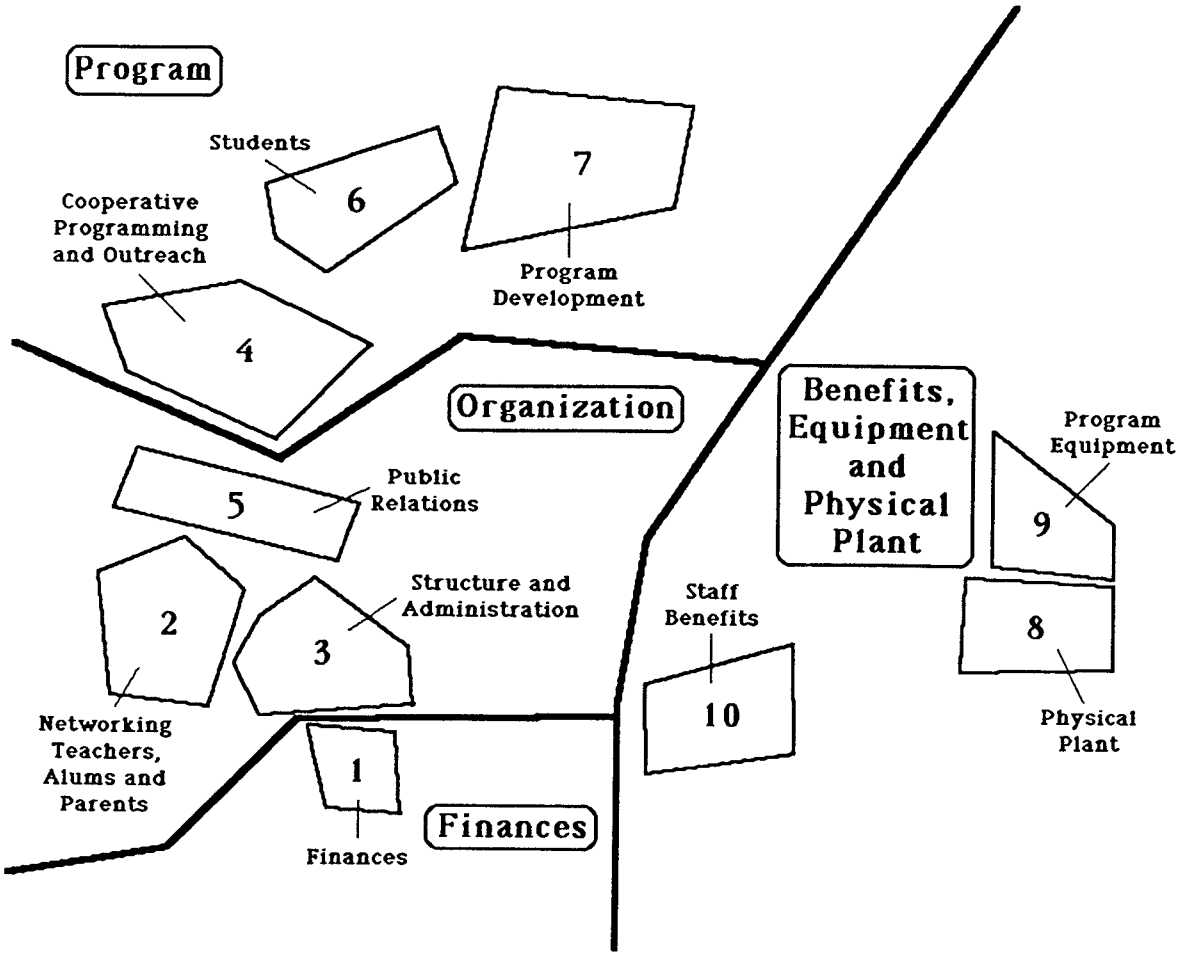


Figure 10. Final concept map for the Community School of Music and Arts Project.

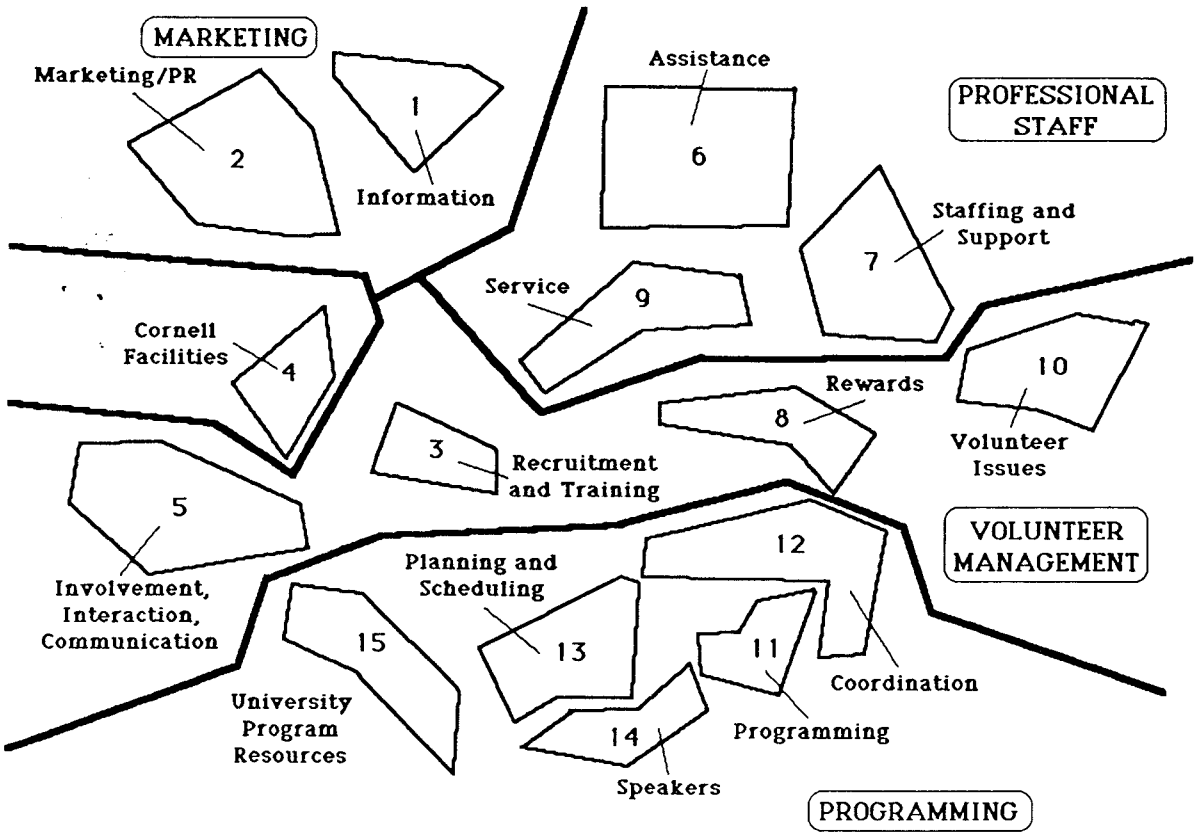


Figure 11. Final concept map for the Alumni Affairs Project.

and 15 clusters. The clusters at the bottom of the map were related to program issues while the others were more concerned with the management of the alumni offices themselves (i.e, marketing, professional staff, and volunteer management). On the basis of this map, the alumni directors could devise new programs and approaches for improving services.

Student Life (Student Assembly)

The student assembly is the central student governance body in a major university. They used concept mapping to help them to identify the issues which they might address during the next academic year. Approximately 20 elected student representatives generated 87 statements which "describe issues in students' lives which the student assembly might address." The final map is shown in Figure 12. One striking feature of this map is the degree to which it resembles the two maps discussed earlier (at least at the regional level) in connection with student life issues (Projects 8 and 9 above). There are four identifiable regions—Academics, Personal Growth, Social Interactions, and University and Community Issues—and 15 clusters. More will be said about similarities between concept maps below.

Employment

Between 8 and 10 persons representing various groups involved in employment participated in this project, including members of the County Board, the Department of Social Services, and various agencies which provide employment services. They generated 90 statements which described "issues related to employment services in the county." The final map is shown in Figure 13. Five regions and 16 clusters were identified. The largest region in the upper left of the map described the needs and issues of the target population. The "context" region in the lower left was related to legislative and funding issues. The region which extended from the center of the map to the lower right referred to interagency coordination—an issue of clear importance to the representatives of the different agencies who participated in this process and compete for clients, resources, and support. The region in the upper right is related to planning such things as employment training and to better understanding the demographics of the working population in that area. Finally, the "service and support" region had to do with the direct services—such as job placement—which various agencies provide.

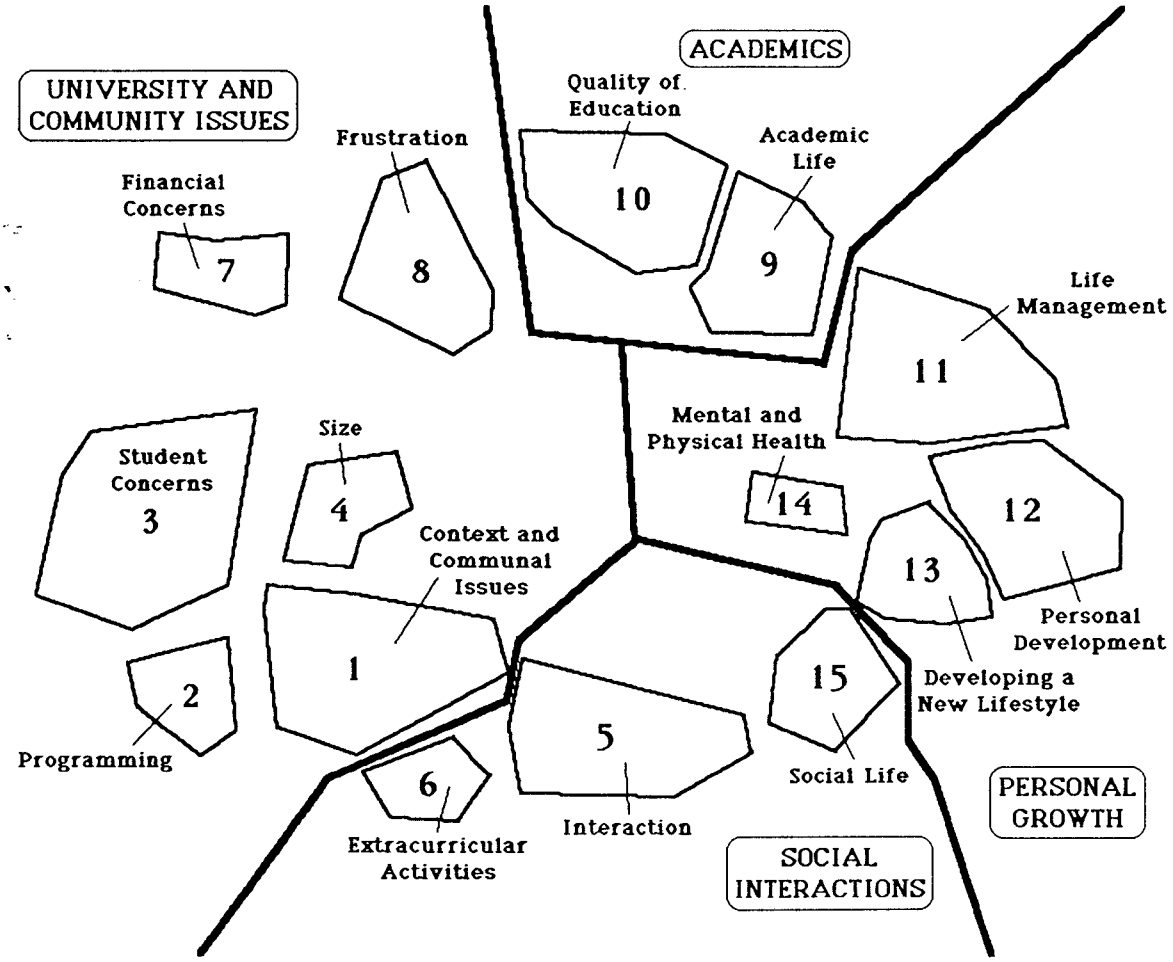


Figure 12. Final concept map for the Student Life (Student Assembly) Project.

Personnel Management

In this project, a four-person department at a water filtration plant wished to examine the current job responsibilities for persons in the department and the need for additional personnel. All four men (including the supervisor) participated in the brainstorming session which yielded 74 statements describing the different tasks which they perform in their jobs. Because there were only four people, each was asked to sort the statements twice. In addition, each was asked to check on a rating sheet any of the tasks which they themselves routinely perform on the job. The final map is shown in Figure 14. There were five major regions and twenty clusters identified. In addition, maps were produced for each of the four individuals (not shown) showing the statements on the map which they had some responsibility for. It was immediately apparent that there were few well-delineated areas of responsibility—each person wound up doing tasks from all over the map. On the basis of this map, the participants decided to do one additional step to assist them in planning. First, they rated each cluster for how many man-hours they needed to accomplish that task ideally in a typical month. Second, they rated (individually and then totally) how many man-hours they currently spend on each task in a typical month. These two ratings could then be graphed onto the map in Figure 14 to provide a visual indicator of where they believe they are most discrepant from the ideal in terms of personnel resources. Assuming that there are discrepancies, they might then reallocate current job responsibilities and/or use that information as justification for an additional personnel request. A mapping of job responsibilities like this one may enable an organization to identify areas where they have over or under-committed personnel and resources.

Counseling Services

A counseling agency which provides a wide range of mental health services wished to develop a conceptual

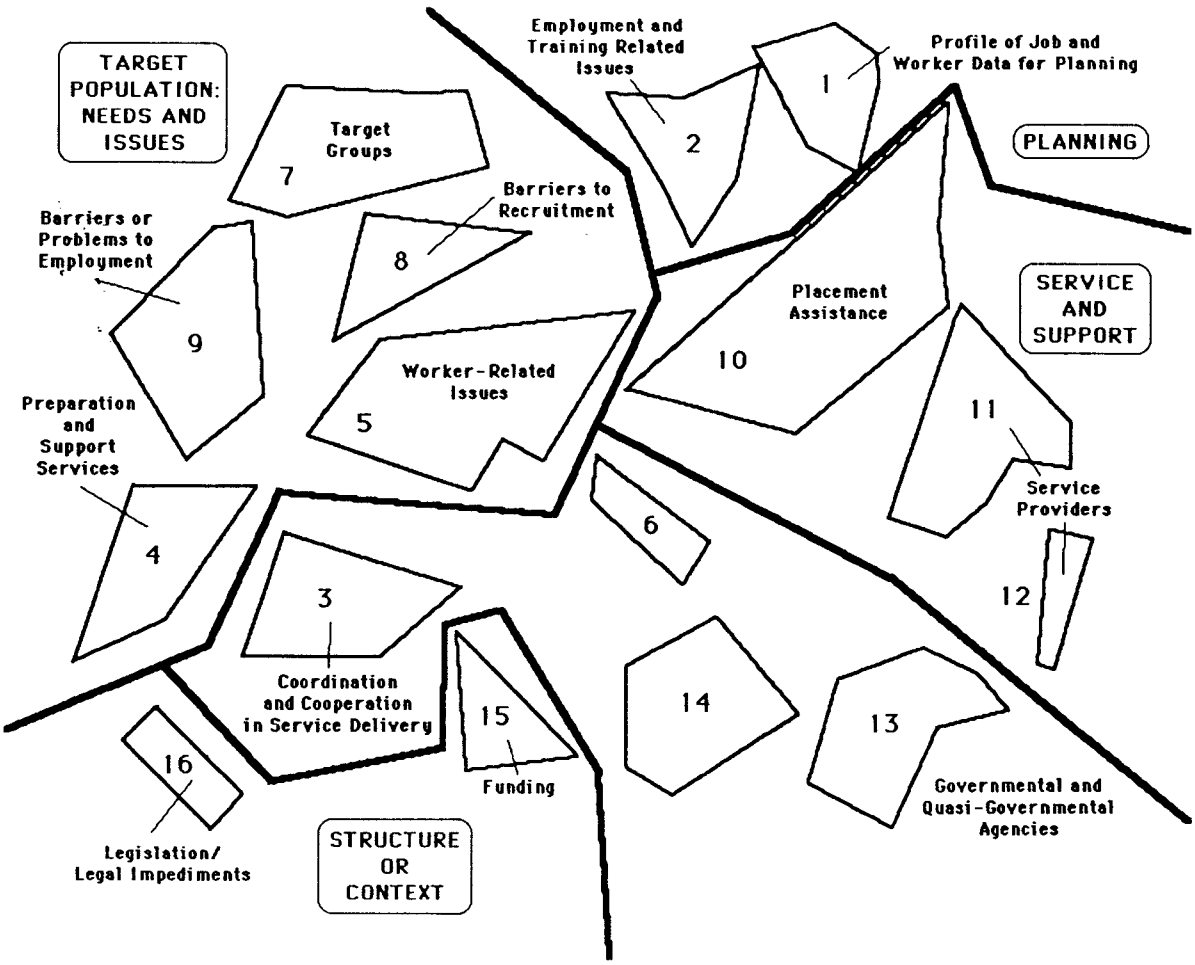


Figure 13. Final concept map for the Employment Project.

framework for its long-term strategic planning effort. A small group consisting of 10–15 key staff members and members of the Board of Directors generated 80 statements which described services which the agency does or might provide. In addition to sorting these statements, each person rated them on a 1 (lowest priority) to 5 (highest priority) scale. The final rating concept map is shown in Figure 15. Four regions and 15 clusters were identified. At the top of the map are clusters related to major mental illness which in the past has been the primary service focus for this organization. On the left are medically related services. Community services are on the bottom and specially targeted outpatient services on the right. The highest priorities were given to the clusters for substance abuse, family oriented outpatient services and services for the elderly. While the first two had already been addressed to some extent by the agency, there was currently no program for the elderly. Consequently, one outcome of this process was the recognition that they wished to consider what services might be needed by the elderly. They decided that the best way to begin to address this task would be to involve other local agencies with responsibilities for the aged to join in a concept mapping process on that topic (see project 17 below). In addition, they intend to use the map as the basis for examining planning data such as budgetary and staff time allocations by cluster, competition from other agencies for services in each cluster, and so on. Just as with the priority ratings, each of these additional variables can be overlaid on the original concept map to provide a visual display of the data.

Senior Citizens

A consortium of organizations which deal with senior citizens wished to use concept mapping to identify the issues which needed to be addressed in their area. Between 10 and 15 staff and Board members participated. They generated statements which represented what they believed needed to be addressed in order to

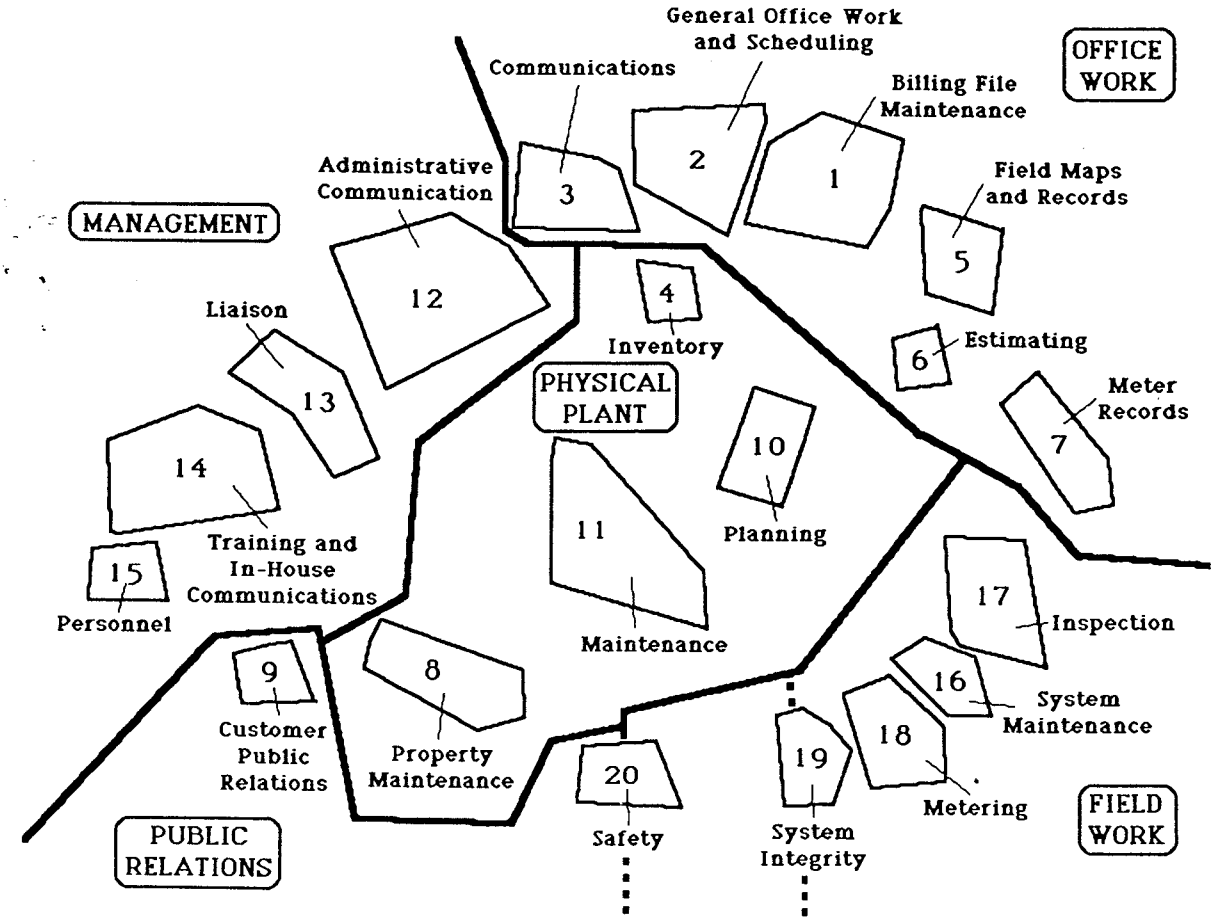


Figure 14. Final concept map for the Personnel Management Project.

improve services to the elderly. The final map is shown in Figure 16. Five regions and 12 clusters were identified. Clearly, clusters on the left of the map (and especially on the lower left) were judged as more important. Thus, training and development, community education, and volunteerism were considered areas which needed special attention, whereas self-help services and recreational activities did not.

Elderly

This project emerged from project number 15 described above. On the basis of that earlier project, an agency identified services for the elderly as a priority for planning. In order to begin addressing this, they assembled a consortium of agencies which deal with the elderly in their county, including representatives of the United Way and of health and mental health services. The 10–15 participants generated 95 statements which described "the issues, problems, concerns or needs which the elderly have in the county." In addition to sorting, each participant rated the statements on a 1 (lowest priority) to 5 (highest priority) scale. The final map is shown in Figure 17. There were 18 clusters identified and the group did not wish to define regions on the map. The most intriguing result of this process was the sense within the group that they had hit upon a "theory" regarding the development of the elderly. They suggested that the map could be interpreted nicely in a counter-clockwise manner beginning with the rightmost cluster, personal growth and education. The "healthy" elderly would view this as a priority (along with socialization) and are also the most likely to be politically active and powerful. Consequently, much of the political power available to the elderly is spent on advocacy for issues for the well elderly. As people grow older (continuing counter-clockwise) they increasingly run into issues related to access to resources (e.g., coordination of local services, costs). Eventually, housing issues become more predominant involving questions about home care, outreach, supervised care, and the potential for abuse. Clearly at this point the family is extremely relevant and likely to be involved in the deci-

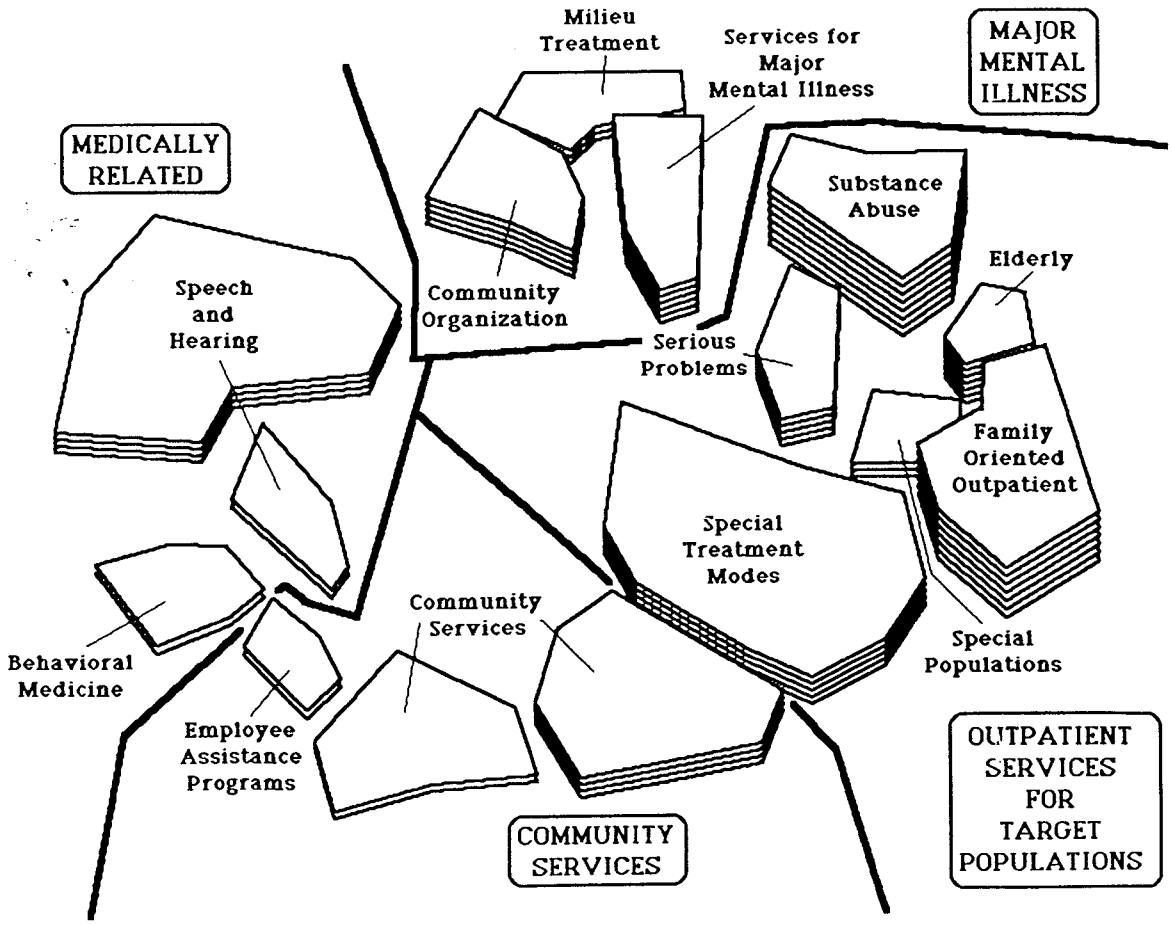


Figure 15. Final concept map for the Counseling Services Project.

sion making. Eventually with advanced age there are issues of competence, vulnerability, depression, fears and helplessness and, ultimately, decreasing functioning and death. This counter-clockwise interpretation led to a stimulating discussion where participants concluded that it was important to intervene early in this cycle and involve the well-elderly as advocates for issues which are on their horizon. In addition, the map is being used as the basis of an examination of current service patterns in the county (by cluster) to determine whether the agency which initiated this project should continue to consider developing services in this area.

Arts Council

An Arts Council which is responsible for fostering and encouraging cultural and artistic efforts in their county wished to use concept mapping as the basis of their long-term planning process. Approximately 10–15 members of the Board of Directors generated 63 statements which described "what should be done by an effective Arts Council." In addition, each participant rated each statement on a 1 (lowest) to 5 (highest) priority scale. The final map which resulted is shown in Figure 18. The Arts Council is a relatively small organization (1 FTE professional, a part-time secretary, and volunteers) which means that it relies on its Board Members more directly to be active in addressing the mission of the organization. Prior to the concept mapping, there was little consensus among Board Members concerning what their roles and functions should be. On the basis of this project, they were able to identify their major tasks as helping to seek funding, encouraging the educational function of the Council, long-term community involvement, and public relations for the Arts Council and the arts in general. In addition, they clearly saw the need for ongoing and expanded Board development efforts.

Planned Parenthood

A Planned Parenthood organization wished to use concept mapping as the basis for long-term planning. They involved a relatively large group of between 30 and 35

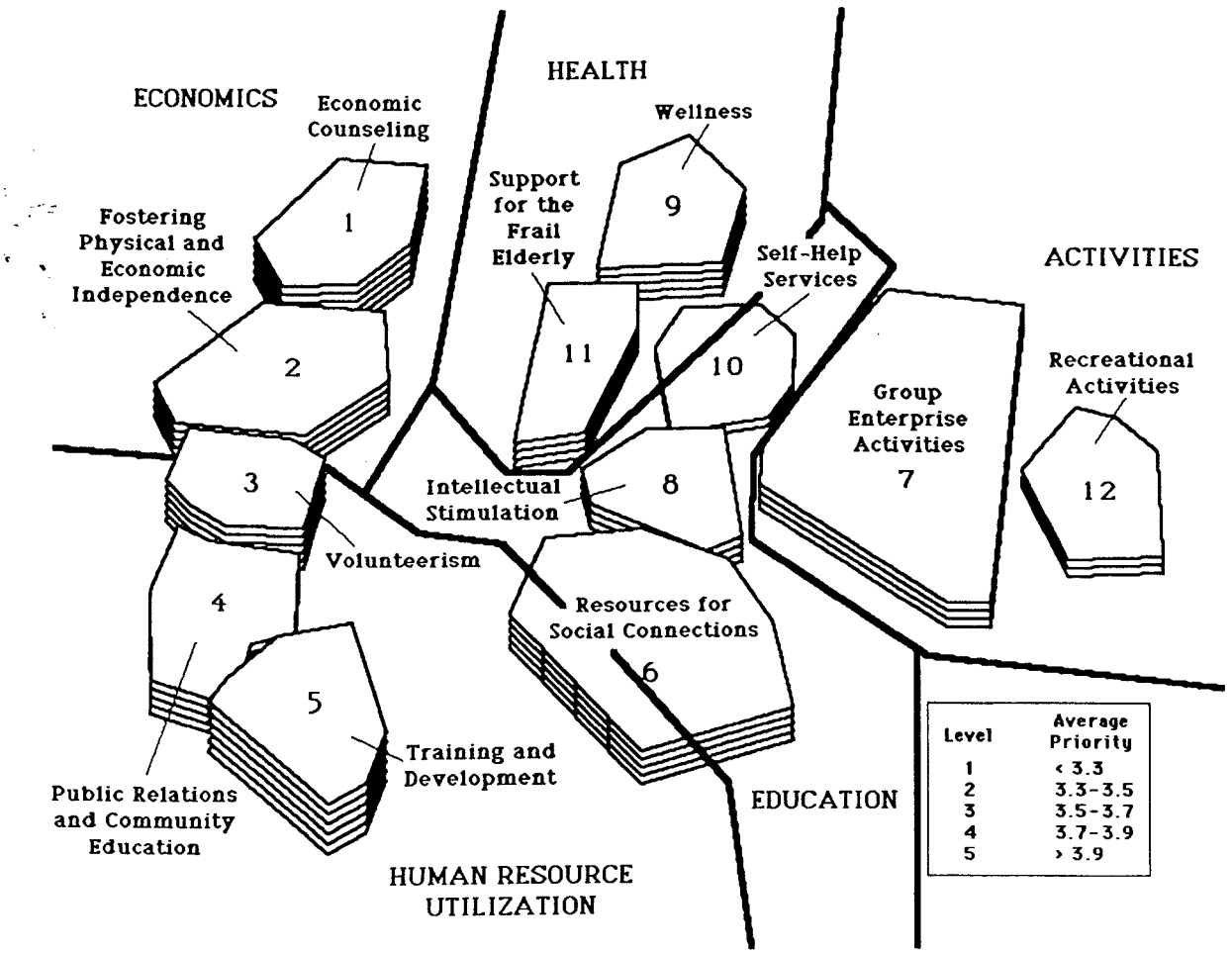


Figure 16. Final concept map for the Senior Citizens Project.

staff and board members in generating 89 statements which described issues which they needed to address in the longer term future. The final map (with priority ratings) is shown in Figure 19. There were 7 regions and 18 clusters identified. In the interpretation, it became apparent that the region pertaining to education on the lower left was given the highest priority, closely followed by financial issues and public relations concerns.

Music and Arts in Daycare

In this project, concept mapping was used to develop the framework for constructing training sessions for daycare providers in music and art activities for preschool children. Between 10 and 15 daycare providers generated 61 statements which described the types of issues which they wished to see addressed in training in music and arts. In addition, each participant rated each statement on a 1 (not at all) to 5 (extremely) importance scale. The final map is shown in Figure 20. Six clusters were identified in the analysis. The highest importance was assigned to teacher training issues, followed by issues related to the skills and attitudes which teachers were expected to have. The clusters and individual statements were used to plan for the workshops which were later administered. Thus, in this case, concept mapping enabled the client group to devise the issue structure for their own training.

CONCEPT MAPPING AS SOFT SCIENCE: SOME LINGERING ISSUES

The initial motivation for the development of the concept mapping process described here was largely scientific. The thought was that concept mapping could help in the articulation of the concepts used in social research and in their translation into operationalizations. From the outset it was important to establish: (a) that

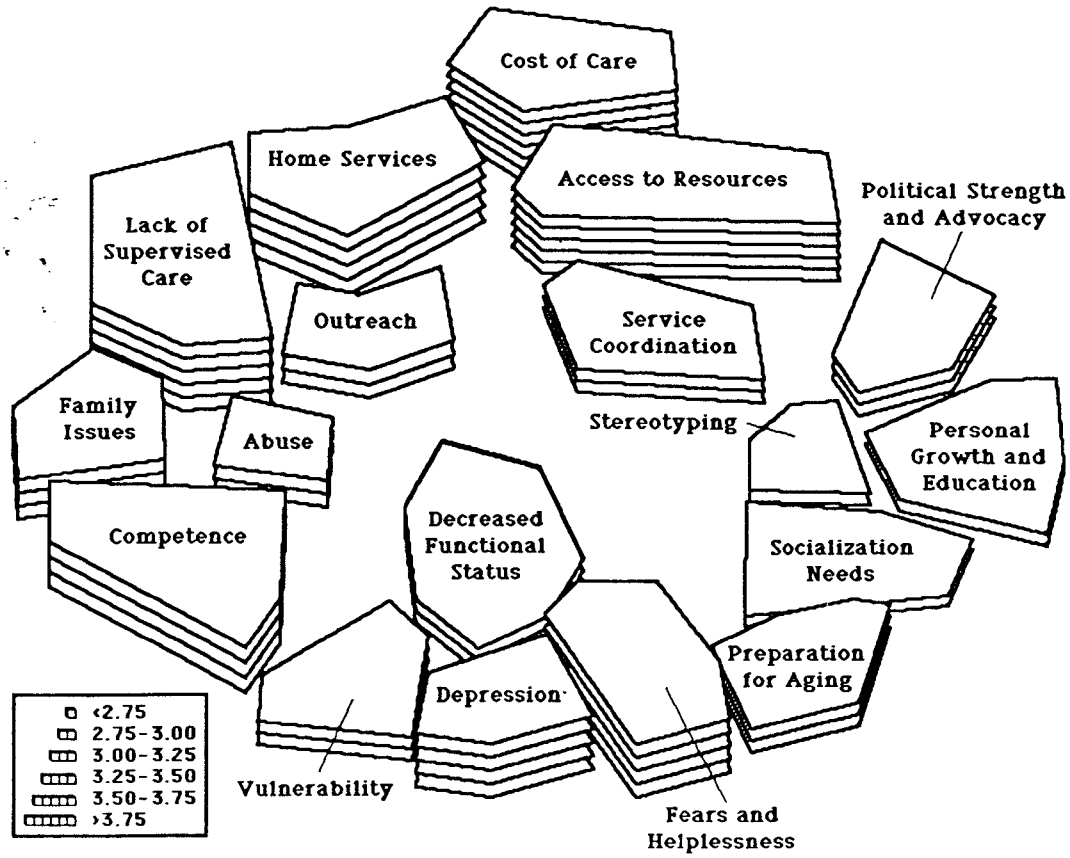


Figure 17. Final concept map for the Elderly Project.

the concept mapping process provided an accurate representation of what people were thinking (i.e., reliability and validity), and; (b) that the concept maps could be integrated into scientific theory-building and experimentation. These two issues are considered separately below. At this point, it is not clear how well the concept mapping process addresses these two issues and so the discussion here will be preliminary in nature and largely suggestive of some of the methodological research which might be undertaken to investigate these concerns further.

Reliability and Validity of Concept Mapping

To date, there have been no major attempts to investigate the reliability and validity of concept mapping. For purposes of discussion, reliability will be understood here to mean the degree to which a map is "repeatable." Validity is meant to refer to the degree to which a map accurately reflects reality.

In terms of reliability, there are a number of questions which could be asked—one could look for overall replicability (e.g., similarity across maps), or look at the reliability of a specific step in the process. A number of studies suggest themselves:

1. Reliability of brainstorming
2. Reliability of sorting
3. Reliability of ratings
4. Reliability of cluster labeling
5. Reliability of final concept maps

For each of these, one could look at the degree to which the same individual or group gets similar results on multiple occasions or the degree to which several equivalent groups (i.e., randomly assigned) independently produce similar results. For instance, we could assess the degree to which we get similar maps when we perform the same process twice on the same group of participants at two different times (a type of test-retest reliability) or we could look at the degree of similarity between maps based on separate processes carried out simultaneously by random subgroups of the same population (actually, a type of convergent validity). Each type of study would pose it's own methodological difficulties. For instance, how would one assess the degree of similarity between two sets of brainstormed statements, two sets of cluster labels, or two maps which were constructed using entirely separate processes? For the reliability of sorting it would be possible to corre-

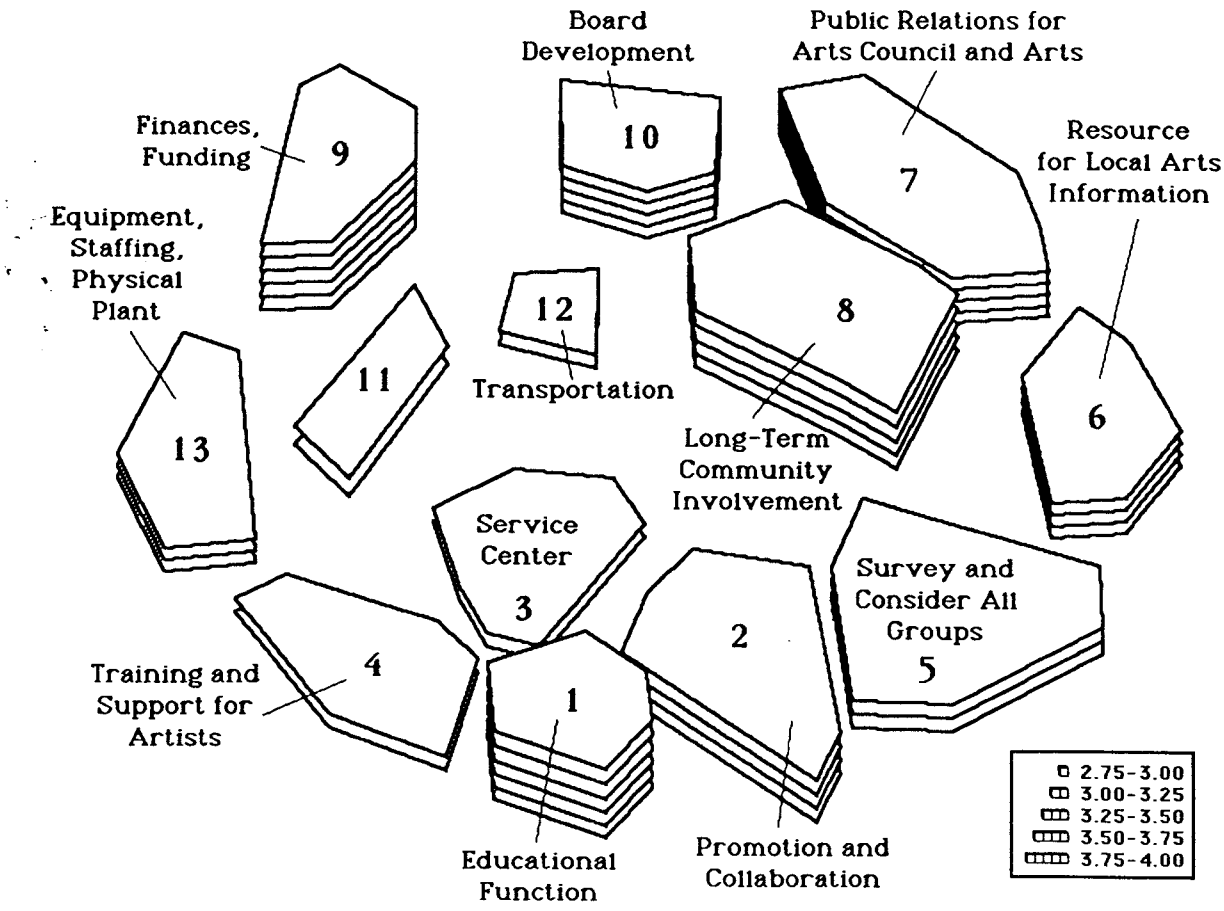


Figure 18. Final concept map for the Arts Council Project.

late participant's binary similarity matrices. Reliability of ratings could be assessed with a simple correlation between the ratings.

While no direct evidence is available on reliability, we can get a rough indication by visually examining maps from similar populations on similar topics. For instance, it is clear from the presentation of projects above that several of the maps which describe issues in students' college lives (Figures 8, 9, and 12) have some striking similarities in features despite the fact that they were the results of three entirely different processes with three very different groups of students at three different times. Or, one could look at the similarity between maps from similar organizations, such as the projects from the Community School or Music and Arts (Figure 10) and the Arts Council (Figure 18); or from similar topic areas such as aging (Figures 16 and 17). In all of these cases, however, the projects differed in important ways which might minimize the degree to which we would expect similarity in concept maps. For instance, the purpose of Project 9 was to have Residence Advisors project what *they* thought the issues were in undergraduates lives whereas in Project 10 the students were asked directly what the issues are. Lack of agreement between maps could thus be attributable either to low reliability or to the different nature of the projects. Clearly, there is a need for research which is explicitly constructed to examine reliability issues.

In terms of validity of concept mapping, the only direct evidence on the question comes from the work undertaken by Dumont (this volume) who looked at the degree to which computed concept maps correlated with hand-placed maps (a type of convergent validity). The evidence is somewhat ambiguous and seems to indicate that estimates of validity depend largely on the level at which one looks for agreement and the manner in which the validity estimate is computed. Although the preliminary results are promising, Dumont's work needs to be replicated with more people, different types of participants, and in different subject areas.

There are several general approaches which might be taken in investigations of the validity of concept mapping. One method would be to compare concept maps (or results of any step in the process) with comparable information generated by some other method, as Dumont (this volume) did when looking at computed

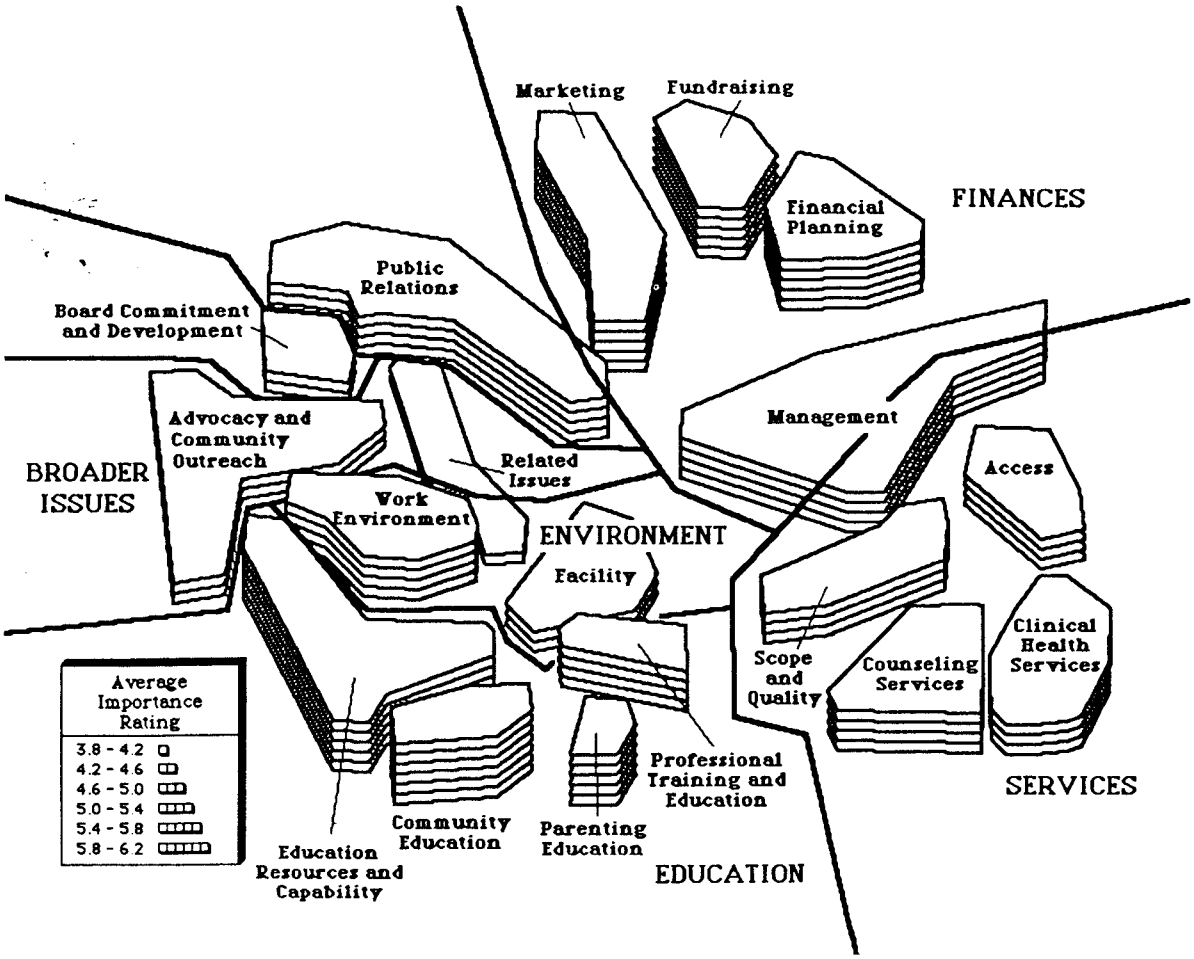


Figure 19. Final concept map for the Planned Parenthood Project.

versus hand-placed maps. For instance, one could compare a set of brainstormed statements with transcripts of interviews on the same topic to see whether similar issues arise.

A second method for examining validity would be to see whether participants could identify the "correct" concept map from a set, much like a witness identifying an accused criminal in a line-up. For instance, let's say that in addition to generating the computed concept map for a project we also generate three more maps which have the same statements on them but where the statements are randomly placed on the map. The validity question is whether the participants could identify the computed map as the one which most accurately reflects their thinking. Incidentally, this type of study, more than almost any other, addresses the distinction between soft science and hard art raised here. If participants cannot distinguish the map computed from their data from randomly generated ones, there would be no effective argument for the validity of this process. On the other hand, if we asked people to tell us which maps were the most suggestive, interesting, or creative, it might very well turn out that randomly-generated ones would be chosen. After all, if people can sensibly interpret randomly generated ink blots in psychological testing, why wouldn't they be able to form interpretations of random concept maps (which would still use the statements which they brainstormed)? In fact, there is some reason to think that deliberate random arrangement of statements on a map—while not meeting the standard of validity (or accuracy in representing what the participants actually think)—might be a good method for getting people to see new relationships and to think creatively.

Finally, it would be possible to examine validity by looking at whether concept maps confirm theoretically expected differences. For instance, we might have two groups of participants—say, teachers and students—involved in a study where we have some clear idea of how we expect these groups to differ in their conceptualizations. Comparison of their concept maps could help to confirm or deny our expectation. Similarly, we

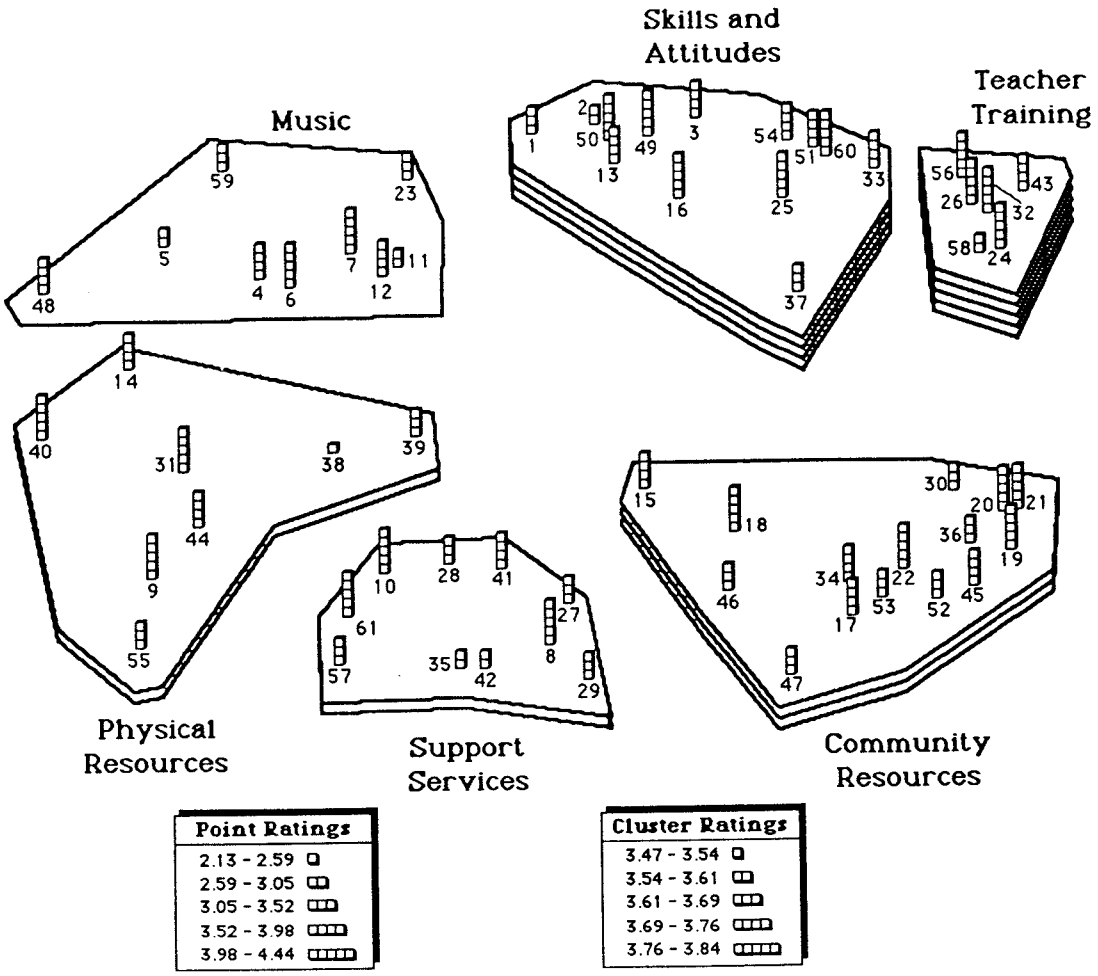


Figure 20. Final concept map for the Music and Arts in Daycare Project.

might have students do a conceptual map at the beginning and end of a course and see whether any changes in the structure of the map correspond to what was communicated in the course or whether the final map is more like the teacher's than the initial one was.

Concept Mapping in Theory Building

How might concept mapping be useful in the social sciences? Consider the plight of the graduate student who needs to define the major constructs for a dissertation project. While all of the texts on research say that it is important to define constructs, there is no concrete advice given on *how* to articulate a conceptual framework. The concept mapping approach views concept definition as a measurement task—much like that of developing a scale. The student generates lots of statements which described the construct(s) in question, and then organizes those statements in some way. Or, consider the lack of conceptual clarity in much of the psychological research which is published—a sentiment well articulated by Sartori (1984) for psychology and the other social sciences. For instance, what are the distinctions/relationships between terms like "self esteem," "self worth," "self image," "locus of control," "dependency," and so on? Or, what are the distinctions/relationships between the terms "intelligence," "achievement," "academic performance," and so on? In addition, over the past few years there has been a growing recognition in the field of evaluation research that improvement is needed in the types of theories which guide evaluation projects. In particular, Bickman (1986) has called for more theory-driven evaluation and Chen and Rossi (1983, 1987) have argued that "grand" social science theory about social programs needs to be augmented by or replaced with more concrete theories about how programs function in real-world settings. All of these considerations suggest that concept mapping might be useful for improving theory-building in the social sciences.

It is important to distinguish between theories and concepts. One thing to recognize is that while theories are built upon concepts, concepts are not, in and of

themselves, theories. The concept maps shown in this paper do not necessarily constitute theories. A theory postulates a relationship—usually causal—between two or more concepts. A concept map provides a framework within which a theory might be stated. Perhaps an example would help to clarify the distinction. Let's assume that we are interested in evaluating a program designed to improve the self esteem of a certain group of high school students. In trying to define self esteem we use concept mapping to describe all of the terms we can think of which are related to our notion of self esteem. Presumably, terms which are more similar in our minds will be closer on the map. In this way we might distinguish several sub-aspects of self esteem and might set self-esteem within a broader framework of other concepts related to the self. However, the concept map itself does not constitute a theory regarding the effect of our program on self esteem. To achieve such a theory we need to state how the independent variable (i.e., the program) is related to the concepts on the map. For instance, after reviewing the program in detail, we might conclude that some aspects of self esteem on the map will most likely be more strongly affected than others. Specifically, we have overlaid our expectations about program effects onto the conceptual structure, showing where it will affect some concepts and not others. Thus, concept maps can act as the framework for a statement of theory, but are usually not considered a theory in and of themselves.

This discussion suggests that concept mapping may be particularly useful for theory-driven social research because of its detailed, visual, pattern-based representation of concepts. Trochim (1985, in press) has described this under the framework of "pattern matching." In pattern matching, one needs a theoretical pattern and an observed one. The theoretical pattern should describe the relationships or outcomes which are expected. The observed pattern consists of the relationships or outcomes which are measured. To the extent that these patterns match and there are no other theories which would account for the observations as well, one can conclude that the theory in question is supported. Pattern matching works best when there is a clearly articulated, detailed theoretical pattern because detailed patterns are more likely to be unique and a match will, consequently, be attributable to this unique theoretical "fingerprint." Concept mapping is particularly valuable for pattern matching because it can help researchers to generate (scale) their theoretical expectations in detail.

Pattern matching is a general approach to research. In evaluation, it can help to guide the development and assessment of the program, sample, measures and outcomes. If concept mapping has utility here, it's value will be far-ranging. The theory of pattern matching is discussed in greater detail in Trochim (1985, Chen & Rossi, 1983, 1987) and is illustrated by the papers in this volume by Davis, Marquart and Caracelli. We have only begun to explore the use of concept mapping for pattern matching methods. For instance, Trochim (1988) shows how concept maps can be used as the framework for exploring patterns in randomized clinical drug trial data. More research is needed on the utility of concept mapping for scientific theory building in general and for pattern matching in particular.

REFERENCES

BICKMAN, L. (Ed.). (1986). Using program theory in evaluation. *New directions for program evaluation*. San Francisco, CA: Jossey-Bass.

CHEN, H.T., & ROSSI, P.H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283-302.

CHEN, H.T., & ROSSI, P.H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10*, 95-103.

GUROWITZ, W.D., TROCHIM, W., & KRAMER, H.C. (1988). A process for planning. *National Association of Student Personnel Administrators Journal, 25*(4), 226-235.

SARTORI, G. (1984). *Social science concepts: A systematic analysis*. Beverly Hills, CA: Sage.

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

TROCHIM, W. (1988). *The effect of Aplrazolam on panic: Patterns across symptoms*. Unpublished manuscript, Cornell University, Ithaca, NY.

TROCHIM, W. (in press). Pattern matching and program theory. *Evaluation and Program Planning*.

TROCHIM, W., & LINTON, R. (1986). Conceptualization for evaluation and planning. *Evaluation and Program Planning, 9*, 289-308.

Evaluation and Program Planning, Vol. 12, p.111, 1989
Printed in the USA.

0149-7189/89 $3.00 + .00

ISSUE CONTRIBUTORS

Valerie Caracelli recently completed her PhD in the Department of Human Development and Family Studies at Cornell University. Her current research represents parallel interests in developmental psychology and program planning and evaluation.

Leslie J. Cooksy is a doctoral candidate in Human Service Studies at Cornell University. She received her M.S. at Cornell in Development Sociology. Her current research focuses on the use of a pattern matching designs to evaluate employment assistance program theory.

James E. Davis is currently Assistant Professor in the Department of Educational Studies at the University of Delaware. He received his PhD in Program Evaluation Studies at Cornell and was a postdoctoral fellow at the Educational Testing Service in Princeton, New Jersey. His research areas have included computer simulations of program evaluations, computer-based evaluation and data collection, and outcomes assessment in higher education.

Jeanne M. Dumont is a PhD candidate in the Department of Human Service Studies at Cornell University. Her previous research has focused on social policy analysis, including the impacts of legal reform on recipients of mental health services. Her current research interests are in mental health evaluation, and program theory and development.

Patrick Galvin is currently a doctoral candidate in the Department of Education at Cornell University. His thesis research examines questions of equity associated with delivery of educational services through the organization of shared service cooperatives like the BOCES in New York State.

Douglas Keith is a PhD candidate in Cognitive and Developmental Psychology in the Department of Human Development and Family Studies at Cornell University. The subject of his thesis concerns visual modeling of abstract concepts. For the past two years, he has also been involved with work in the field of artificial intelligence.

Rhoda Linton is an independent program evaluator and a part-time core faculty member of the Union Graduate School. She received an MSW from Syracuse University and a PhD in Program Evaluation Studies at Cornell. Her current research focuses on the documentation and analysis of women's organizing.

Marc Mannes is Senior Program Manager at the American Indian Law Center, Inc., Director of Project INTACT, and a PhD candidate in Human Service Studies at Cornell University. His dissertation research focuses on planning for the implementation of social technologies.

Jules M. Marquart has a PhD in Program Evaluation from Cornell University. Her research interests include conceptualization and measurement issues in evaluation, and employer-provided family support systems as well as other work and family issues. She is currently using the method described herein to develop and validate a child care assessment instrument for the government of Singapore.

William M.K. Trochim is Associate Professor in Program Evaluations Studies in the Department of Human Service Studies at Cornell University. He has conducted research and published in such areas as: quasi-experimental research designs (especially the regression-discontinuity design); computer simulations for program evaluation; implementation and quality control of evaluations; conceptualization methods for evaluation and planning; and pattern matching approaches to inference making.

Kathleen Valentine received an MS in Psychiatric Mental Health from Syracuse University and is currently a PhD candidate in the Department of Human Service Studies at Cornell. Because of her years of clinical management experiences in nursing she has developed a research focus in health care evaluation. More specifically, she has an interest in the use of evaluation techniques to enlighten decision makers about the need to include such "intangibles" as caring in policy and reimbursement issues rather than excluding such variables because of problems in quantifying them.