of doing that, I had to go back and start questioning some of these earlier decisions, which irritated the hell out of the National Academy of Sciences, who were offering me the job of evaluating their new 40 million dollar project. And so I didn't get it and deserved not to get it. I mean, I was misbehaving. So these are things that you learn from experience. You learn your own strengths and weaknesses. But for me, that's a heavy problem. I can be 10 times more valuable than a service evaluator but only to somebody who is willing to put up with the fact that I'm going to kick down the walls of the box and there's not that many people who will do that. National Academy of Sciences is a very conservative body, and they have made gross mistakes which I could have just shut up about and maybe eventually got them to remedy the mistakes later on, once they had some trust in me, or maybe simply saved up until I wrote the final report and then pointed all these mistakes out, but no, I couldn't shut about up it. I had to tell them there were mistakes right away. There we go, 30,000 bucks a year for 5 years. There it is.

Jane: Of all your various contributions to evaluation, what do you see as most significant?

Michael: The concept of evaluation as a transdiscipline.[7]

Jane: So, of all your contributions to the field, which are the most misunderstood or insufficiently well grasped?

Michael: Well, goal free is the one that people really got pathetic about. It sort of brings out the worst in people. . . . At any rate, so goal free is probably the thing that is most misunderstood.

Jane: It seems there is a big proportion of the field that knows about your work by what they read somebody else said about it.

Michael: There's a lot of that when it comes to the nature of transdiscipline: I need to write more books!

Jane: So, in terms of getting people clear on the concepts that you think are really important to push the discipline forward, what do you think we've learned from how evaluation has evolved so far that we should try to apply to make sure we get on the right track?

Michael: Perhaps the key point is not to forget those outside our own discipline. Changes come from external pressure as well as internal. Well, I think it is no different from any other discipline. The birth of disciplines is always like this. The history of evolutionary theory, the history of molecular biology, the history of signal theory, the history of string theory, there is loads of this stuff going on in their early days. But I think the obligation really is on me to make it simple and with some help from you doing the same thing, we may be able to get enough out there that swamps the misinterpretations. . . . I don't care about their delay so much . . . my main interest is to get it right. . . . to get the ideas straight. So, I do a lot of teaching, workshops, et cetera, but I've still got so many things that I haven't got straight yet, that that's going to keep me busy for a while!

Notes

1. Editor's Note: For another personal account of his professional life, refer to Scriven, M. (2004). Reflections. In M. C. Alkin (Ed.), *Evaluation roots* (pp. 183-195). Thousand Oaks, CA: Sage.

2. Tyler, R. W., Gagne, R. M., & Scriven, M. (Eds.). (1967). Perspectives of curriculum evaluation. *American Educational Research Association Monograph Series in Curriculum Evaluation, 1*(1, Whole No. 1).

3. Editor's Note: The first meeting was held in Snow Mass, Colorado, in the summer of 1973.

4. Editor's Note: See Patton (1997) for a brief description of early professional groups for evaluators. Patton, M. Q. 1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

5. Editor's Note: Michael refers to the Paul F. Lazarsfeld Award for Contributions to Evaluation Theory and the Alva nd Gunnar Myrdal Awards for Evaluation in Government Service and Cumulative Contributions to Evaluation 'ractice.

6. Scriven, M. (1991). *Evaluation thesaurus.* Newbury Park, CA: Sage.

7. Scriven, M. (2003). Evaluation in the new millennium: The transdisciplinary vision. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems* (pp. 19-41). Mahwah, NJ: Lawrence Erlbaum.

Method Notes

This section includes shorter papers (e.g., 10-15 double-spaced manuscript pages or less) describing methods and techniques that can improve evaluation practice. Method notes may include reports of new evaluation tools, products, and/or services that are useful for practicing evaluators. Alternatively, they may describe new uses of existing tools. Also appropriate for this section are user-friendly guidelines for the proper use of conventional tools and methods, particularly for those that are commonly misused in practice.

Concept Mapping as a Technique for Program Theory Development

An Illustration Using Family Support Programs

Scott R. Rosas

Center for Children's Health Innovation

Abstract: The emergence of theory-driven evaluation has drawn attention to the need for practical ways for evaluators to construct program theory to achieve the benefits of the approach. This article illustrates the use of concept mapping as a technique to explicate underlying program theory in the context of family support programs. In this example, 29 professional staff members from a statewide family support program brainstormed 96 specific benefits for participant families. Fourteen direct service staff members sorted and rated the items and interpreted the map produced by multidimensional scaling analysis and hierarchical cluster analysis. Some of the benefits of the technique for evaluators conducting theory-driven evaluations are highlighted. These include helping evaluators decide what to measure, capturing the complexity of programs in the context of implementation, and managing good stakeholder-evaluator relations. The general limitations and issues of concept mapping as it applies to theory development are also discussed.

Keywords: *concept mapping; program theory; family support; outcome framework*

The growth and popularity of theory-driven evaluation has increased in recent years. Viewed as an alternative to traditional method-driven approaches, theory-driven evaluation has become a part of mainstream evaluation practice, and several exemplars demonstrating the benefits exist (Donaldson, 2003). Theory-driven evaluation has been credited with improving program conceptualizations, supporting sound implementation, creating value for stakeholders, and improving evaluation design sensitivity (Birkmayer & Weiss, 2000; Donaldson, 2003;

Scott R. Rosas, Center for Children's Health Innovation, Nemours Health and Prevention Services, Christiana Building, Suite 200, 252 Chapman Rd., Newark, DE 19701; phone: (302) 444-9131; fax: (302) 444-9200; e-mail: srosas@nemours.org.

Author's Note: I thank Mel Mark for his insightful comments and guidance in the preparation of this article.

American Journal of Evaluation, Vol. 26 No. 3, September 2005 389-401
DOI: 10.1177/1098214005278760

Hacsi, 2000; Huebner, 2000). Key challenges for theory-driven evaluation have been the shortage of practical advice for evaluators, ways to manage stakeholder relations, and confusion over what is meant by program theory (Donaldson, Gooler, & Scriven, 2002; Weiss, 1997).

A critical requirement of theory-driven evaluation is the development and articulation of a clear program theory. Leeuw (2003) recently suggested that methods for constructing program theory could be organized into the policy-scientific, strategic assessment, and elicitation approaches. Sources of information used in theory specification have included stakeholder input from facilitated group work (Chen, 2005), program documentation, in-depth interviews, and observations (Bickman, 1985; Wholey, 1987), social science concepts, and previous research in the program domain (Donaldson, 2001; Riggin, 1990). Some have argued for the integration of methods and sources in the construction of program theory because of the inherent strengths and weaknesses found in each (Bickman & Peterson, 1990; Riggin, 1990).

One technique that can contribute to the specification of program theory is concept mapping (Trochim, 1989). Concept mapping is a multistep process that helps articulate and delineate concepts and their interrelationships through group process (brainstorming, sorting, rating), multivariate statistical analyses (multidimensional scaling [MDS], hierarchical cluster analysis), and group interpretation of the conceptual maps produced. The technique's foundation is rooted in cognitive and organizational psychology. Concept mapping can be viewed as a variant of the elicitation approach for constructing underlying program theory (Leeuw, 2003). Caracelli and Riggin (1994) reported that concept mapping has demonstrated potential in improving evaluation by providing a systematic mechanism for articulating and visually representing interrelationships of key concepts. Indeed, concept mapping has been used across a number of human service settings in a variety of ways to support evaluation (Trochim, 1986).

Examples of the wide utility of concept mapping have been published recently in the evaluation literature. Yampolskaya, Nesman, Hernandez, and Koch (2004) demonstrated how concept mapping could help develop specific components of logic models and articulate a theory of change. Shern, Trochim, and LaComb (1999) used concept mapping to assess fidelity when program models were transferred from program designers to staff members. Similarly, Barth (2004) used concept mapping to identify dimensions for rating program quality from which actual data gathered during site visits could be compared. Despite its benefits, including structured participant input and explicit criteria for distinguishing between concepts, concept mapping is used less often than might be expected. Few examples exist in the evaluation literature of concept mapping being used with the specific intent of explicating underlying program theory in theory-driven evaluations (Marquart, 1990; Mercier, Piat, Peladeau, & Dagenais, 2000; Trochim & Cook, 1992).

Family support programs provide a valuable opportunity for illustrating how concept mapping might be used for program theory development. Models of family support programming are relatively new and have emerged from a variety of theoretical and nontheoretical approaches (Gottlieb, 1988). Models have included the use of support networks to mediate stress, prevent psychological and emotional problems, promote child and family development, and empower individuals and families (Bond, 1982; Cochran, 1991, 1992; Elias & Branden, 1988). However, specification has been lacking as to which of these (or other) elements are implemented and in what ways. In particular, questions remain as to how empowerment is conceptualized as a specific part of a family support approach. This conceptual ambiguity has been a long-standing issue, and theory-driven evaluations have been limited to date. Indeed, Olds (1988) indicated that the lack of a clear theoretical framework has left evaluators and researchers unable to specify the benefits families might receive from family support initiatives. This gap has led to confusion about what to assess and to the propensity to measure an unwieldy number of outcomes in the child and family environment (Weiss, 1988). Consequently, evaluating outcomes of family support programs has proved to be a challenge (Powell, 1987; Family Support America, 2001; Weiss & Jacobs, 1988).

To further illustrate the benefits of concept mapping for evaluators, this article presents an application of the technique in which the views of the benefits expected from a set family support programs were solicited from program staff members. It was expected that this process would lead to the development of a theoretical outcome framework that would enhance a subsequent theory-driven evaluation.

Method

Participants

Participants in the concept-mapping exercise were professional staff members from various family support program sites throughout the state of Delaware. This included individuals responsible for providing direct services, as well as those serving in various administrative roles, including supervision, quality assurance, and site management. Participants were purposefully selected because of their direct connection to the processes and outcomes of the program. Overall, 80% of participants were female. Participants represented geographically diverse regions, with program sites located in both rural and metropolitan areas. More than 80% of the participants were African American, 10% were Hispanic, and 7% were Caucasian. Participants averaged 35 months of involvement with the program, but the range was considerable (e.g., three administrators reported 10 or more years of involvement).

Procedure

Concept mapping was used to build the framework of intended benefits of the family support model as currently implemented. The concept-mapping process was accomplished in six separate sessions, each session lasting between 1 and 2 hours. The Concept System, a proprietary software package designed for the concept-mapping process, was used to handle the entry, processing, analysis, and reporting of the data.

Brainstorming sessions. The first three sessions were conducted with a total of 29 participants. The three brainstorming sessions were held on different days and in separate locations. The brainstorming session activities were included on the agenda at regional quarterly staff meetings. Eleven staff members participated in the first session, 10 in the second session, and 8 in the third session.

The context, rationale, and process for the inquiry were described at the beginning of each session. It was explained that the primary intent was to obtain participants' ideas and perceptions of the family support program of which they were a part. Following the introduction and background, the data collection process began with the brainstorming of items guided by the focus statement: "Generate statements which describe the specific benefits that family members engaged in the family support program should experience." Items were listed on flip charts as articulated by participants and clarified only when needed. A running list was made at each session. Following all three brainstorming sessions, the lists were compiled, items eliminated on the basis of redundancy, and a final statement list was drafted. The generated items served as the core content for the sorting and rating procedures. This data collection method produced a total of 94 items, which are displayed in Table 1.

Table 1
Brainstormed Items With Mean Importance Ratings Grouped by Cluster

Item Number	Statement	*M*	*SD*
Life skills cluster			
2	Family activities occur without drugs or alcohol	4.93	0.27
44	Opportunities to expand knowledge of ATOD prevention	4.71	0.73
12	Opportunities to develop and improve their communication skills	4.50	0.85
65	Increased exposure to educational opportunities	4.43	0.51
92	Opportunities to learn how to budget and control finances	4.14	0.86
94	Opportunities to learn how to create resources	4.07	0.73
1	Opportunities to develop and improve money management skills	3.86	0.95
32	Opportunities to develop and improve time management skills	3.86	0.53
4	Increased work productivity	3.36	1.01
	Average rating		4.21
Resource control cluster			
66	Increased exposure to employment opportunities	4.21	0.70
57	Ability to access resources	4.14	0.86
86	Opportunities to develop relationship building skills	4.14	0.77
8	Increase knowledge of resources available in the community	4.00	0.78
67	Increased exposure to home ownership opportunities	4.00	0.88
76	Opportunities to learn how to access support	3.93	0.83
53	Opportunities to learn how to establish fun family activities	3.79	0.97
38	Increased exposure to new and different activities	3.57	1.16
17	Opportunities to have different experiences	3.43	0.94
	Average rating		3.91
Social connectedness cluster			
31	Opportunities to develop a sense of leadership	4.29	0.73
50	Opportunities to develop and improve social skills	4.07	0.83
74	Socialization with other families occurs	3.93	0.73
87	Opportunities to experience support from group members	3.93	0.92
5	Opportunities to learn how to deal with others outside of the family	3.86	0.86
29	Opportunities to experience a sense of belonging to a group	3.86	1.03
18	Decreased feeling of isolation	3.71	0.99
20	Increased peer support	3.57	1.02
	Average rating		3.90
Community connectedness cluster			
7	Increased ability to ask for help from others	4.07	0.92
54	Sense of community pride is fostered	4.00	0.88
47	Opportunities to organize a group and effect change	3.93	0.92
83	Opportunities to become community leaders	3.93	1.00
85	Increase volunteerism in programs	3.93	0.92
45	Opportunities to contribute to community	3.86	1.03
71	Opportunities to develop ownership in the program	3.86	0.86
84	Opportunities to build trust with professionals	3.79	0.70
11	Increased ability to work in a group	3.71	0.73
51	Increased exposure to cultural activities	3.71	0.73
77	Opportunities to participate in community events	3.71	0.83
35	Working relationships with professionals are developed	3.64	0.74
24	Opportunities to learn how to form partnerships	3.50	0.94
91	Increased understanding of the local political system	3.29	0.99
	Average rating		3.78
Family guidance cluster			
70	Positive parent role models for family members	4.93	0.27
81	Increased family responsibility from parents	4.79	0.43
62	Parents actively work to end family dysfunction	4.71	0.47
68	Bonding occurs among family members	4.71	0.47
56	Parents help foster a sense of family trust	4.57	0.76
15	Parents facilitate increased family closeness	4.50	0.65

(continued)

Table 1 (continued)

Item Number	Statement	*M*	*SD*
69	Goals set for families by parents	4.50	0.52
3	Parents foster a more family oriented attitude	4.43	0.65
73	Opportunities to improve interpersonal skills between family members	4.29	0.73
42	Opportunities to develop trust between family members	4.21	0.89
48	Respect is fostered among family members	4.21	0.80
	Average rating		4.53
Personal growth cluster			
80	Parents take responsibility for themselves	4.71	0.61
52	A sense of self-esteem is developed by parents	4.64	0.63
30	Increased self-esteem for parents	4.57	0.65
72	A sense of independence is developed by parents	4.57	0.65
78	A sense of self-worth is developed by parents	4.57	0.94
27	A sense of self is developed by parents	4.43	0.65
6	Parents develop a sense of trust	4.36	0.84
25	Ownership of responsibilities by parents occurs	4.29	0.73
79	A sense of appreciation is developed by parents	4.29	0.99
19	Increased confidence by parents	4.21	0.89
22	Increased life satisfaction by parents	4.21	0.89
55	Parents foster family closeness though activities	4.14	0.77
23	Decreased sense of shame by parents	4.07	0.92
	Average rating		4.39
Critical reflection cluster			
75	Healthy lifestyle is developed by parents	4.71	0.47
93	Improved adult life skills	4.64	0.50
26	Ability to make decisions is increased	4.50	0.52
28	Parents understand their role in the family	4.50	0.85
46	Problem-solving abilities are developed and enhanced	4.43	0.65
63	Parents gain knowledge of how to change their family structure	4.43	0.65
16	Better decisions are made by parents	4.36	0.74
21	Goal setting behavior is improved	4.36	0.93
61	Family dysfunction and impact is recognized	4.36	1.15
58	Sense of opportunity and possibility is enhanced	4.21	1.05
82	Parents engage in parenting classes	4.21	0.89
33	Planning skills are developed and enhanced	4.00	0.68
	Average rating		4.39
Parenting behaviors cluster			
14	Closer relationships with children are built	4.86	0.36
90	Parent-child communication is improved	4.86	0.36
10	Overall parenting skills are improved	4.79	0.43
89	Children are helped to make better choices	4.71	0.47
13	Listening skills are improved	4.64	0.50
34	Children are held accountable for their actions by parents	4.57	0.94
40	An awareness of children's needs is developed	4.57	0.76
43	Youth development is supported	4.57	0.76
49	Children are taught respect by their parents	4.57	0.65
9	Knowledge of child development is enhanced	4.50	0.52
88	Understanding what children face daily is enhanced	4.50	0.76
60	Knowledge of how children cope and for what	4.43	0.76
37	Parents discover new ways to educate their children	4.36	0.63
59	Parents develop a sense of control in their child's life	4.36	0.74
64	Parents learn to advocate for their children	4.36	0.63
36	New ways to discipline children are taught	4.29	0.73
41	Understanding how negative behaviors of children are influenced	4.29	0.83
39	An awareness of children's wants is enhanced	3.79	0.89
	Average rating		4.50

NOTE: ATOD = alchol, tobacco, and other drugs.

Structuring sessions. A smaller group of 14 direct service staff members were reconvened for two separate structuring sessions of the concept-mapping process. Structuring sessions were designed to articulate the interrelationships among the concepts by having each participant sort and rate each of the 94 items. Again, these sessions coincided with regular regional staff meetings. Eight participated in the fourth session, and 6 participated in the fifth session. This subgroup of 14 participants was thought to be the most familiar with the daily activities and families involved in the program. In the earlier brainstorming sessions, the administrative subgroup had been included to ensure that an adequate range of items was obtained. However, it was decided to have those only involved in the day-to-day direct service work participate in the sorting and rating of the items at the structuring sessions.

At the two structuring sessions, participants completed two tasks: a sorting task and a rating task. For the sorting task (Rosenberg & Kim, 1975), the 14 participants received a set of 94 cards, each card with one of the previously brainstormed items. Next, session participants were asked to sort cards into piles according to their similarity. Each worked individually to group the cards into piles "in a way that made sense to them." Thus, participants were asked to decide which items, in their opinions, belonged together. The purpose was to obtain data on how participants perceived the relationships among the items generated in the brainstorming sessions. There were four major restrictions to the sorting process: (a) All items could not be put into a single pile, (b) all items could not be put into their own separate piles (although some items could be grouped by themselves), (c) items could not be placed in two piles simultaneously, and (d) there could not be any "miscellaneous" piles.

For the rating activity, the 14 participants in the two structuring sessions were given the entire list of brainstormed items set in a questionnaire format and asked to rate the items using a 5-point, Likert-type response scale. Each participant was asked to rate the items in terms of how important each item was to his or her idea of program benefits for family members. The response options for the rating scale were *relatively unimportant* (1), *somewhat important* (2), *moderately important* (3), *very important* (4), and *extremely important* (5). Because it seemed unlikely that many of the items were totally unimportant with respect to the program, it was emphasized that item ratings should be considered relative to the importance of the other items in the brainstormed set.

Following the two structuring sessions, MDS and hierarchical cluster analyses were conducted and findings were formatted for use in the interpretive sessions.

Brief Description of MDS

A review of the sort data submitted by the 14 participants revealed it to be complete, and all sorts were included in the analysis. Nonmetric MDS was conducted using the sorted data to produce the concept map. The Concept System uses MDS to create a map of points that represents the set of brainstormed statements, on the basis of the group similarity matrix that resulted from the sorting procedure. MDS is based on the measurement model that assumes that the relative similarity of objects can be represented in terms of the relative distance between pairs of points (Kruskal, 1964).

The sort data from each of the 14 participants were entered into a matrix with 94 rows and 94 columns, corresponding to the 94 brainstormed items. A value of 1 was assigned to a cell for those items sorted together in a pile (e.g., if Items 2 and 26 were sorted into the same pile by Participant A, in Participant A's matrix, 1 would be entered in the cell in which column 2 and row 26 intersected). A value of 0 was assigned to a cell for those items that had not been sorted together. Because an item was considered sorted with itself, the diagonal values of the matrix were equal to 1. The result was a binary symmetric similarity matrix for each of the 14 individuals. The individual sort matrices were added together to produce a group similarity matrix. This matrix contained the same 94 rows and columns; however, instead of ones and zeros in the cells, the group similarity matrix included the number of participants who sorted the items together in a pile. Hence, values between 0 and 14 were possible. A value of 14 indicated that all participants saw the 2 items in question as belonging together. Likewise, a value of 0 indicated that none of the participants saw the items as belonging together. Finally, the MDS analysis transformed the group similarity matrix into a rank-ordered table of distances between items. This table of similarities (or distances) was used to iteratively place points on a map so that the original table was as fairly represented as possible. The output consisted of a set of plotted X-Y values that formed a "point map," with each numbered point representing an item.

Brief Description of Hierarchical Cluster Analysis

Hierarchical cluster analysis was the second analysis conducted using The Concept System software. This analysis was used to group individual statements on the map into clusters of items that presumably reflected similar concepts. The approach used by The Concept System used the two-dimensional X-Y coordinate data obtained from the MDS analysis as input for the hierarchical cluster analysis and applied Ward's algorithm. In effect, this technique grouped or separated the items on the map, as they were placed by MDS, such that items placed in the same cluster were in contiguous areas of the map. The end product was a "cluster map," which revealed how the MDS points were grouped.

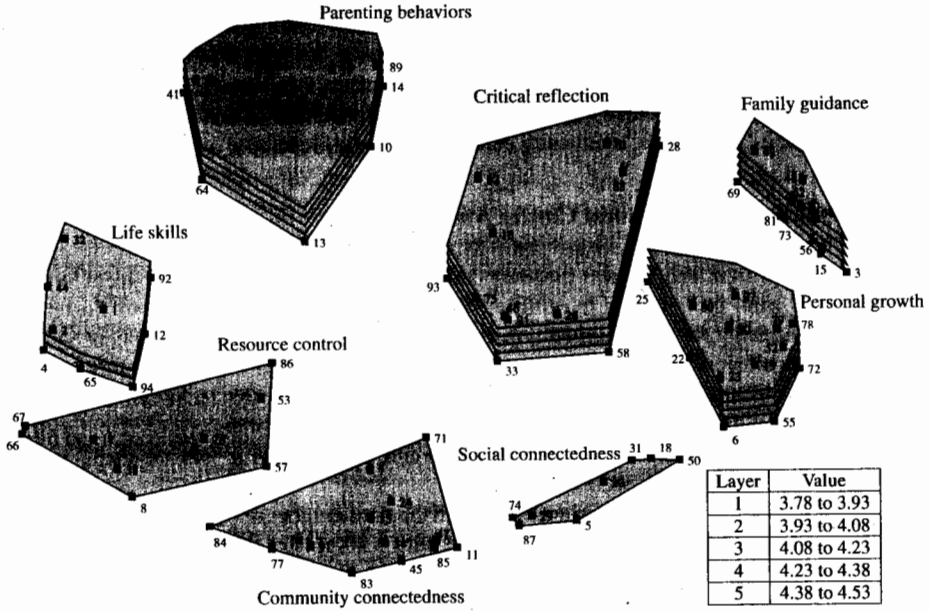
There are no specific, definitive criteria by which the final number of clusters can be selected. The procedure used in this example was to begin with a 10-cluster solution and successively increase and decrease the solutions by 1. The higher and lower solutions were examined and a decision made for each configuration as to whether the separation or merger of clusters appeared to adequately represent the data as stated, organized, and prioritized by participants. From a review of the different configurations, an 8-cluster solution was determined to be the best in preserving the most detail while offering the most substantive interpretation.

A review of the rating data submitted by the 14 participants revealed it to be complete, and the entire set was included in the analysis. The rating data were averaged across the 14 participants for each item and cluster. The average ratings were overlaid on the point map to produce two types of importance rating output. A "point rating map" consisted of the original point map with the average rating per item displayed as vertical columns next to each point. A "cluster rating map" consisted of the cluster average ratings displayed as layers of each cluster.

Interpretive session. The subgroup of 14 participants was reconvened for a final session. The output produced by The Concept System was supplied to each participant for the purpose of labeling clusters as well as offering any preliminary interpretative comments. Participants individually reviewed the brainstormed list, the point map showing the MDS placement of the brainstormed items and their identifying numbers, the cluster map showing the eight-cluster solution, the point map with the average statement ratings overlaid, and the cluster map with the average cluster ratings overlaid. Following the review, a group discussion was held, with participants working cluster by cluster to recommend acceptable labels that captured the content of each cluster. Agreement of the labels among the participants was sought, and in those rare situations in which consensus was not immediately reached, a possible label was suggested.

Following the labeling of the clusters, a guided tour of the map was provided, describing how the analysis constructed the map as well as the meaning of the proximal location of the points. Participants were invited to discuss their interpretations and suggest how the map might be used in the future.

Figure 1
Final Concept Map Depicting Eight Clusters
With Layers Indicating Average Ratings of Importance



Layer	Value
1	3.78 to 3.93
2	3.93 to 4.08
3	4.08 to 4.23
4	4.23 to 4.38
5	4.38 to 4.53

Results

The MDS analysis of the similarity matrix converged after four iterations, producing a final stress value of .04 (Kruskal's Stress Formula 1) and $R^2 = .99$. The stress value measured how well the particular configuration represented the data. It is essentially a goodness-of-fit measure, with lower values reflecting a stronger relationship between the optimal and actual configurations (Kruskal, 1964). In this case, the values provided evidence of a strong fit between the configuration (map) and the actual data (group similarity matrix). Thus, the accuracy of the map in representing the way the group of participants organized the items was confirmed.

The result of the two-dimensional solution of the MDS analysis was a sophisticated map of concepts representing several domains. The two-dimensional configuration of the 94 brainstormed items is graphed in Figure 1. The distances among the points and clusters are fixed in MDS. However, the directionality of the map is subjective, and the map could be rotated in any direction without adjusting the distances. There is no substantive meaning to the fact that "Parenting Behaviors" is at the top of the map and "Social Connectedness" is at the bottom. In the figure, each item is marked by a point with the item identifying number located next to it (for the item corresponding to a number, see Table 1). Layers for each cluster indicate the average importance ratings across all items in that particular cluster. That is, the more layers shown, the higher participants rated the importance of the items in a cluster. Averages represented by the layers are actually double averages, averaging across all the participants and across all the items in a cluster. Therefore, even slight differences in clusters are considered meaningful (Trochim, 1989). In this map, each cluster is shown with its agreed-on label. Location on the map is a function of perceived similarity. That is, the items that were most frequently sorted together are closer to one another than those that were seldom or never sorted together. Clusters of items are indicated by the shaded polygons in Figure 1. As with items, clusters that are closer together are considered to be more similar.

Reliability and Consistency of Concept Maps

The reliability or consistency of the results of the concept-mapping process was assessed through three methods suggested by Trochim (1993). First, to estimate a coefficient analogous to split-half reliability, the sample of 14 participants was randomly divided into two equal groups of 7. Similarity matrices were constructed for each group, and separate MDS configurations were calculated. The two halves were correlated, and the Spearman-Brown correction was applied to estimate reliability. The reliability estimate for the similarity matrices was .83 ($df = 4{,}464$, $p < .001$), indicating statistical consistency in the structure of the split-samples configurations. Second, to estimate a coefficient analogous to average interrater reliability, each participant's individual binary sort matrix was correlated with the total similarity matrix. The 14 correlations were averaged, and the Spearman-Brown correction was applied. The average interrater reliability estimate for the participants' sort data was .94 ($df = 4{,}464$, $p < .001$), again indicating statistical consistency in the sorted relationships across participants. Finally, the internal consistency of the importance ratings was calculated across all 94 items of the scale. Cronbach's α for the importance ratings was .94. Taken together, these results indicate that the output of the concept-mapping process was reliable.

Discussion

The application of concept mapping with this set of family support programs illustrates the potential value of the technique for evaluators conducting theory-driven evaluations. First and foremost, concept mapping can be used to help improve design sensitivity in theory-driven evaluation. Output from concept mapping could help determine evaluation questions; inform choices about design, measurement, and analysis; and support interpretation of results. More specifically, results such as those reported here can help evaluators make decisions about outcome measurement and design a data collection strategy that matches the intended benefits of the program under study (Trochim & Linton, 1986). In this example, the ratings of importance documented which aspects of the program's theoretical outcome framework are most salient to program practitioners and likely to be susceptible to their influence. This information can be used to compare the program's current measurement strategy to determine the degree to which it is consistent with what practitioners perceive as important. Often, program staff members are convinced that their programs have impacts that are not represented in the assessment batteries of most evaluations (Larner, 1992). Thus, concept mapping can help evaluators facilitate the alignment of what program practitioners think should happen (theory) and what actually does happen (observation), thereby enhancing claims of construct validity.

Second, concept mapping can improve program conceptualizations by representing the complex relationships found in most programs and thus minimizing the risk for simplification, rigidity, linear thinking, and insufficient analysis.[1] Concept mapping can be used to differentiate "theories in use" from "espoused theories" and help move toward a realistic depiction of the program (Argyris, 1982). Indeed, variations in underlying program theories, including differences in intervention targets, goals, and expectations, have revealed implications for evaluation

and the success of programs (e.g., Duggan et al., 2000; Powell, 1988). In the current illustration, the concept map highlighted the multiple relationships between staff-perceived benefits and by their relative location on the map indicated the strength of those relationships. Thus, evaluators are able to recognize that in practice, outcomes are not viewed independently but are related in a multiple domain framework. Furthermore, the identification of an explicit focus on empowerment within the family support program is one example of the conceptual benefits that can arise from the use of concept mapping. In comparison with the cursory guidance on empowering parents found in the program materials, the level of detail offered by program practitioners as to the expected benefits for program participants was fairly sophisticated. The concept map depicted a rich description of psychological empowerment, as it might exist in practice, which is conceptually consistent with previous theoretical and empirical work on the subject (cf. Zimmerman, 1995). The presence of a focus on empowerment within the model is an important issue for any theory-driven evaluation of the family support program. Clearly, practitioners perceive that program participants should experience benefits related to empowerment as a result of their engagement in the program, even in the face of limited information about how this might be facilitated and achieved. This should signal to the evaluator the need to consider how empowerment is operationalized in the implementation of the model that might not be obvious in the program documentation.

Finally, concept mapping can be a valuable, practical tool in building and maintaining good stakeholder-evaluator relations. In general, theory-driven evaluations require considerable interaction with stakeholders to construct the program theory, determine evaluation questions, and solicit input for the design. Proponents of theory-driven evaluations have argued for the need to recognize the operator-dependent nature of programs and account for the assumptions and perceptions of those most closely associated with the program (Rossi, 1978; Weiss, 2000). As illustrated in this example, concept mapping facilitates active involvement in the creation of knowledge whereby the evaluator is not isolated from the thinking of program staff members about what occurs during implementation. It uses the language of those engaged in the process rather than the terms of the evaluator. Indeed, the process of concept mapping can be as important as the product, because stakeholder input is elicited and incorporated throughout the entire process. Buy-in from program staff members can be beneficial to evaluators conducting theory-driven evaluations. Huebner (2000) found that staff members tended to be more cooperative throughout the course of an evaluation when it was conducted on the basis of a common understanding and shared view.

Concept mapping can help translate complex conceptual ideas into easily understood visual representations, forming the basis of program theory, as illustrated in the present example. A strength of the technique is that concept mapping is designed to bring order to the task of theorizing, which is often difficult for groups to accomplish in less structured approaches. Concept mapping blends qualitative and quantitative methods to elicit, represent, interpret, and assess the quality of a variety of program conceptualizations. It could be used, for example, to articulate specific elements related to program activities as an additional step in the development of a comprehensive program theory. Thus, a map of the intended benefits (outcomes) could be combined with a map of program activities (processes) to offer a more complete program theory.

Limitations and Issues of Concept Mapping

Concept mapping is one technique for use in theory construction; as such, it should be viewed as only one source of information when building comprehensive program theory. It should be noted that concept maps themselves do not necessarily constitute theory. Rather, they represent a conceptual framework and provide a foundation from which a program theory might be derived. In addition, the specific individuals who participate limit the breadth and depth of the conceptualizations that emerge from the concept-mapping process. Riggin (1990) emphasized that theorizing about a program occurs at different levels, depending on the knowledge base and experience of those engaged in the process. Although inclusion of widely diverse groups may be important for the validity of conceptualizations, this may also make program theory specification more difficult and less specific (Marquart, 1988). Therefore, careful attention should be given to which stakeholders are involved in the concept-mapping process and what their relationships are to the program under study. To protect against misspecification through unchecked acceptance of stakeholder viewpoints as valid program theory, Chen and Rossi (1983) argued for the verification of stakeholder conceptualizations with available social science concepts.

Accuracy in program theory specification also depends on and is limited by the scope of the brainstorming prompt. Variations in the prompt would ultimately lead to different conceptualizations, because different content would emerge during brainstorming. Thus, clarity during the brainstorming phase is critical to avoid incorrect interpretations. For example, in this illustration, the prompt focused on the generation of perceived benefits. It is important to recognize that what emerged from this process was theorizing about outcomes, not a complete program theory. Thus, it would be inaccurate to claim the concept map represents a more comprehensive program theory without further work.

Finally, concept mapping requires either the use of specialized, proprietary software or a thorough understanding of multivariate analyses and graphical representation. Although the available software package enables users to easily manage all steps of the concept-mapping process, the cost may be prohibitive in theory-driven evaluation efforts for which resources are limited. Conversely, all phases of the process can be completed manually using standard word-processing, statistical, and graphics applications. However, the time required by the evaluator and commitment from the participants may not be reasonable in many situations.

Conclusion

Chen and Rossi (1983, 1987) argued for more concrete theories about how programs work. More recently, emphasis has been given to methods and techniques that can make program theory construction transparent and more precise (Leeuw, 2003). Without a detailed theory or model, the benefits of theory-driven evaluation are difficult to achieve. Concept mapping can offer evaluators an additional technique to include in their repertoire as they consider ways to specify program theory as part of theory-driven evaluations.

Notes

1. For a detailed discussion of the issues related to the use of ecological models in family support evaluation, see Weiss (1988).

References

Argyris, C. (1982). *Reasoning, learning, and action: Individual and organizational.* San Francisco: Jossey-Bass.

Barth, M. C. (2004). A low-cost, post hoc method to rate overall site quality in a multi-site demonstration. *American Journal of Evaluation, 25,* 79-97.

Bickman, L. (1985). Improving established statewide programs: A component theory of evaluation. *Evaluation Review, 9,* 189-208.