All social research is based on pattern matching ideas. A pattern match involves a correspondence between a theoretical or conceptual expectation pattern and an observed or measured pattern. Two quasi-experimental designs—the nonequivalent dependent variable design and the reversed treatment design—illustrate pattern matching logic well. In program evaluation three pattern matches are important: the program pattern match that assesses program implementation; the measurement pattern match that assesses the validity of the measures; and the effect pattern match that assesses the causal hypothesis. Conceptualization methods are needed to facilitate the articulation of rich theoretical patterns. An example of a conceptualization study is presented and the utility of conceptualization methods for pattern-matching research is discussed.

PATTERN MATCHING, VALIDITY, AND CONCEPTUALIZATION IN PROGRAM EVALUATION

WILLIAM M.K. TROCHIM
Cornell University

All social research is based on the relationship between the ideal and the real, theory and observation, the conceptual and the operational. Typically there is a theory, however well articulated, and the research essentially consists of an attempt to determine the degree to which observations correspond to or "fit" this theory. This may involve an examination of how discernible the hypothesized statistical model is in the variables that were measured, or how well the case notes, field records, or observers' impressions fit some theoretical category scheme. We can describe this central research task generally as the *pattern matching* (Campbell, 1966) of conceptual and operational domains.

AUTHOR'S NOTE: *I wish to thank Rhoda Linton, Chuck Nocera, and Will Shadish for their thoughtful comments on previous drafts of this article.*

EVALUATION REVIEW, Vol. 9 No. 5, October 1985 575-604

This pattern-matching view of social research is strongly related to developments of the notion of construct validity over the past thirty years (Cronbach and Meehl, 1955; Campbell and Stanley, 1963; Campbell and Fiske, 1959; Cook and Campbell, 1979). Construct validity refers to the degree to which observationalizations can be said to reflect their theoretical constucts. Construct validity depends on pattern matching. In order to establish, for instance, that measures reflect the constructs that they are supposed to reflect, we need to have (1) a conceptualization or theory of the expected relationships between the constructs of interest and related constructs from which they must be distinguished (the conceptual pattern); (2) observed interdependencies between purported measures of the constructs of interest and of related constructs (the operational pattern); and (3) a "match" between these two patterns.

This article is concerned primarily with the development of pattern-matching strategies that are appropriate for testing the effects of social programs. Here, the interest is in determining whether the program in question can be said to have *caused* some specific effect. Internal validity concerns the degree to which such causal claims are legitimate for the immediate context of the evaluation. We can claim that we have internal validity (Campbell and Stanley, 1963; Cook and Campbell, 1979) when for the specific context of the study, the program is the most plausible explanation for an observed pattern of effects.

This article seeks to make several points. First, attempts to improve and extend pattern-matching logic will directly enhance our ability to establish construct validity. In addition, we will be in a stronger position to make causal assertions about social programs (i.e., have greater internal validity) if we do a better job of articulating the theoretical outcome patterns that we expect in our research. The importance of pattern making for causal hypothesis testing can be seen clearly through examining two lesser-known, underutilized research designs presented in Cook and Campbell (1979). Second, the development of pattern expectations is viewed as a problem of conceptualization. Although traditional social science theory can and should help to *inform* the evaluation conceptualization, it will seldom by itself be *sufficient* for accomplishing this task well in evaluation contexts. Third, this task of discovering and articulating the conceptual patterns can be viewed as a *methodological* problem. We need to seek out and develop conceptualization methods that go beyond traditional logical-analytic approaches and can improve our ability to generate conceptual patterns. Because conceptualization is a ubiquitous problem, almost every discipline has developed strategies or techniques that are candidates for a taxonomy of conceptualization methods. Some of these are mentioned later and one is illustrated in some detail. Finally, we need to improve greatly our ability to conduct pattern-matching studies. Both the conceptual and statistical issues are complex and severely tax the capabilities of current methodological technology.

PATTERN MATCHING FOR CAUSAL HYPOTHESIS TESTING

In most program evaluations the implicit conceptual patterns are relatively simplistic. Chen and Rossi (1984) are among recent critics who point out, for instance, that the experimental paradigm for evaluation research assumes that the treatment or program is a "black box" or unitary entity, the effects of which can be assessed by determining whether the pattern "program versus no-program" is detectable in one or more outcome measures. The best evaluations, of course, go beyond this simple pattern by using multiple implementations of the program, investigating the degree of program implementation, incorporating multiple measures and measurement methods, and so on. Even in these cases, the emphasis has been more on "having multiples" than on articulating the pattern of interrelationships between these multiples. For instance, it is one thing to have even multiple measures of three major effect constructs (e.g., immediate knowledge recall measures, attitudinal measures, and behavioral measures) and quite another to specify that the largest effects are expected on immediate recall, more moderate effects on attitudes, and the smallest effects on behavioral indicators. It is this latter sense of specifying the hierarchical dimension or continuum across causal or effect concepts that best characterizes the pattern matching ideas discussed here.

In this section, two quasi-experimental research designs that are especially dependent upon pattern-matching ideas are examined. From this discussion, some general principles underlying pattern matching causal hypothesis testing studies are identified.

TWO SIMPLE PATTERN MATCHING DESIGNS

In their discussion of "generally interpretable" quasi-experimental designs, Cook and Campbell (1979) offered, somewhat tentatively, two designs that illustrate well the relationships between internal validity, construct validity, and pattern-matching ideas. Indeed, it is the contention here that the ability to infer cause from these designs is especially dependent on pattern-matching ideas.

The nonequivalent dependent variable design. This design requires pre- and postprogram measurement of at least two variables. The design rests on two principles: (1) there are theoretical reasons to believe the program will affect some of the measures and not others; and (2) the measures are conceptually similar and likely to be affected in the same way by other factors irrelevant to the program. A simple example is offered for the hypothetical case of the evaluation of a special math education program designed to improve scores in geometry. If one measures ability in geometry as well as abilities in other mathematical areas (e.g., fractions, additions), presumably the program should only affect the former and not the latter. If pre-post change is detected in the geometry scores and not on the others, one may conclude that the change may be due to the program and not to other factors, assuming that all variables were measured well. Clearly, there are still other explanations for the program effect which remain plausible. If, for instance, the geometry training increased student interest in studying math more generally, then the students might study and improve in fractions or addition, thus masking the comparison. Nevertheless, to the extent that the measures are conceptually similar, some major inferential problems can be ruled out. For example, if it is reasonable to assume that in the absence of the geometry training students would mature at equal rates on all measures, the obtaining of a differential gains supports the plausibility that the geometry training and not differential maturation rates is the more plausible causal agent. If instead of measures of ability with fractions or addition, we used measures of knowledge of history or current political events, it would be less reasonable to assume that these measures would—in the absence of any geometry training—evidence the same maturational pattern as the math scores. Consequently, if one had to choose, it would be better to use nongeometry math scores as the most immediately applicable "control variables" than to use less related indicators. The nonequivalent dependent variables design has received little use, one exception being a study described by Shadish (1980) that compared "nonverbal" and "verbal" clinical groups and showed that, as predicted, traditional self-report measures showed less differentiation between these two interventions than did behavioral measures of group cohesion.

This simple version of the nonequivalent dependent variables design can be enhanced greatly when we predict more complex patterns rather than that some measures will be affected and others will not. As Cook and Campbell (1979: 120) state:

> We would like now to set the nonequivalent dependent variable design into a broader "pattern-matching" context. The design is obviously strongest where differentiated patterns of change are predicted that allow many alternative interpretations to be ruled out. The probability of ruling out threats depends in part on the specificity of the predicted data pattern so that interpretability increases (1) with the number of dependent variables for which predictions are made—the two variable/two wave case is merely the simplest example of the more general design—and (2) with the specificity of numerical or sign predictions made.

The implication is that the design is best used when a fairly complex theoretical framework can be articulated allowing not just for specification of which variables will be affected or not, but more precisely the degrees to which a set of variables will be differentially affected. If the predicted effect pattern is obtained, one will be in a stronger position to claim that the causal hypothesis is plausible.

The reversed treatment nonequivalent group design with pretest and posttest. The simple version of this design requires two groups that are pre- and posttested on the same variables and are both program groups. The key to the design is that the programs the two groups are in must be conceptually "opposing" such that one is the conceptual "reverse" of the other. This condition means that we would theoretically predict that the two program groups will evidence opposing gain patterns on a single measure of effect—one program will increase scores on the measure while the other will decrease them. The example that is offered (Cook and Campbell, 1979) involves two decision-making procedures as the programs where one is "democratic" or participative in nature and the other is more "authoritarian" or hierarchically controlled. The prediction is that both productivity and job satisfaction will be increased in the *democratic* program group and decreased in the *authoritarian* one.

There are two important notions in this design. First, the programs must carefully be specified conceptually:

> The reversed treatment design with nonequivalent groups is stronger than the no-treatment control group design with respect to construct validity. This is because the theoretical causal variable has to be rigorously specified if a test is made that depends on one version of the cause affecting one group one way and another group the other way (Cook and Campbell, 1979: 125).

Second, the predicted opposite outcome pattern must be obtained. Cook and Campbell (1979) state that "the potentially high construct validity of the reversed-treatment design depends on the research revealing changes in opposite directions." Internal validity also presupposes this pattern match but additionally requires that there are no plausible alternative factors (other than the programs) that could have caused the opposite directions of effects.

The simple version of this design can also be extended to a more general pattern-matching idea. If, in addition to the conceptually opposite programs we include ones that vary by degrees between these poles, we can predict program effects that also vary between the gains expected from one program extreme and the losses expected from the other. Note how more finely varied pattern predictions (if corroborated by the data) enhance *both* construct and internal validity. Construct validity is improved because the causal variable is even more rigorously and richly specified than in the two-program bipolar case. Internal validity is strengthened because with more pattern specificity it is generally less likely that plausible alternative explanations for the observed effect pattern will be forthcoming.

Summary of pattern-matching design issues. There are several important pattern matches that are relevant in causal hypothesis testing. First, there is the *program pattern match.* This involves looking at the correspondence between the program as theoretically conceived and the program as it is operationalized. It is important to note, as Chen and Rossi (1984) argue, that the theoretical conception of the program should include thoughts about how the program will be delivered, the structure of the delivery organization, the facilities and materials used, the relevant client groups, and so on. The program pattern match looks at the degree to which the program was implemented as intended.

Second, there is the *measurement pattern match*. This involves first constructing the conceptual pattern for the constructs to be measured. This pattern is then matched to the pattern of relationships obtained between measures in order to assess whether the constructs of interest are being reflected and distinguished from each other. This can be accomplished by using the multitrait-multimethod matrix approach to examine discriminant and convergent validity (Campbell and Fiske, 1959).

Finally, there is the effect or *outcome pattern match*. On the conceptual side, this involves specifying how the constructs are expected to be affected by the program. For instance, the theoretical effect pattern could consist of the ranking of expected gains across a set of measures. The relationship between this ranking and the obtained gains would constitute the match.

The three pattern matches address different, but related issues, and would normally involve separate analyses in an evaluation. The first analysis would examine program implementation, the second would explore the construct validity of the measures, and the last would assess the causal hypothesis. Nevertheless, it is important to recognize that there is a symmetry or complementarity that exists between the three pattern matches. In general, greater conceptual specificity on any of the three theoretical patterns will lead to increased specifiability of the other two.

To see this, consider first the relationship between the theoretical patterns for the program and the measures. The constructs that are measured in a study should, of course, have some specifiable connection to what the program is doing. If the program consists of training in geometry, then certainly the construct "knowledge of geometry" must be included in the constructs to be measured. More specificity is obtained if the program emphasizes certain types of geometry problems over others. In this case, subconstructs of geometry knowledge that more or less address the emphasized problems could be included in the constructs to be measured. Similarly, if we recognize that conceptually the program in geometry is similar to (but distinguishable from) training in logic, we may incorporate the construct of "logical ability" into the set of constructs to be measured. The implication here is that greater specificity of the program conceptualization will usually imply increased specificity of the measurement conceptualization. If the more specific conceptual patterns can be detected in the data, this implies that improving program implementation (i.e., construct validity of the cause) may have the side benefit of leading to improved construct validity of the measures.

Consider next the relationship between the theoretical program and outcome patterns. Finer distinctions in the conception of the program often will imply hierarchies or continua of expected effects. Thus, when we move from viewing the program as simply geometry training to a more specific statement about what type and how much of each type of geometry problem is emphasized, we imply automatically that there is a hierarchy of expected effects that ranges from more to less emphasized areas of geometry. This idea can be extended when we recognize that there may be some continua of greater conceptual distance as we move from geometry to other mathematical areas (e.g., algebra, elementary mathematical operations), and even beyond to logical and analytic ability, ability in language, and so on. To the extent that such (possibly multidimensional) continua can be specified, they imply a gradually declining pattern of effects that would theoretically be expected. The implication here is that greater specificity in the conceptualization of the program will usually increase the specificity of expected outcomes. Because more complex effect patterns, if corroborated, will usually diminish the likelihood of developing plausible alternative explanations for the complex effects, the implication is that improving program implementation can have the added benefit of increasing the internal validity of the study.

Finally, consider the relationship between the theoretical measurement and effect patterns. Greater specificity about which constructs are to be measured and about the conceptual distinctions between constructs increases the discriminant and convergent validity of the measures. Given a study where knowledge of geometry, logical ability, and language are all relevant, we can predict that geometry measures should be more highly correlated with each other than they will be to measures of logical ability. But if we also believe that the construct of geometry knowledge is more related to logical ability than to language skills, we can predict further that geometry-logic correlations should be higher than geometry-language ones. The crucial point is that this declining pattern of similarity across constructs will often also imply a declining pattern of effects that are expected. Thus, for the geometry program, one would expect the highest effects on geometry measures, lower effects on logical ability, and even lower effects on language.

Note, however, that if the program consists of training in language, the same theoretical pattern would be appropriate for the measures, but the effect pattern would be in the opposite direction (i.e., language → logical ability → geometry). The implication is that although the measurement and effects patterns are distinguishable, improving the construct validity of the measures can have the additional benefit of improving the specifiability of the effects and, consequently, of increasing internal validity.

The discussion of the nonequivalent dependent variables and reversed treatment designs indicates that there are different types of patterns that may be useful. For instance, the simple nonequivalent dependent variables design assumes that the program is a unitary entity and emphasizes a hierarchy of predicted effects *across* effect constructs. In contrast, the reversed treatment design emphasizes a continuum of program variations and looks for a discriminating effect pattern on a *single* effect construct. Obviously, these notions can be combined and would enhance the overall quality of either design. For instance, instead of relying on a single effect construct that is differentially affected by the programs in a reversed treatment design, one could have several discriminating effect constructs and even some specification of which will more or less discriminate the program. Although pattern matching can enable valid causal inference even in the absence of comparison groups (e.g., the nonequivalent dependent variable design) or random assignment to program (e.g., the reversed treatment design), the coupling of such design features with pattern matching will likely further enhance the validity of causal inference.

Pattern matching emphasizes the importance of setting theoretical ideas within some larger conceptual context. This context is essential for identifying what is being studied and distinguishing it from conceptually related ideas that are less germane to the study. Campbell (1966:82) provides an excellent illustration of this idea:

> Imagine the task of identifying "the same" dot of ink in two newspaper prints of the same photograph. The task is impossible if the photographs are examined by exposing only one dot at a time. It becomes more possible the larger the area of each print exposed. Insofar as any certainty in the identification of a single particle is achieved, it is because a prior identification of the whole has been achieved. Rather than the identification of the whole being achieved through the firm establishment of particles, the reverse is the case, the complex being more certainly known than the elements, neither, of course, being known incorrigibly.

The central tension in pattern matching, then, is between the desirability of increased pattern complexity versus the difficulty of detecting or identifying such complex patterns through observation or measurement.

THE ROLE OF THEORY IN PATTERN CONCEPTUALIZATION

If pattern matching is useful for program evaluation, it is important to examine how theoretical patterns can be developed and specified. We can begin such an examination by considering the advice Cook and Campbell (1979: 60-61) offer for determining construct validity (i.e., a pattern match) in causal hypothesis testing contexts:

> First, a test should be made of the extent to which the independent variables alter what they are meant to alter.
>
> Second, a test should be conducted to assess whether an independent variable does not vary with measures of related but different constructs.
>
> Third, the proposed dependent variables should tap into the factors they are meant to measure.
>
> And fourth, the dependent variables should not be dominated by irrelevant factors that make them more or less than was intended.

The critical point is that all of these conditions *presuppose* that a well-conceived set of expectations already exists (as evidenced in the use of phrases like "what they are meant to alter," "meant to measure," and "more or less than was intended"). What is implied here is not a set of simple expectations. Rather, before we can assess construct validity in this way we need a coherent, interrelated framework of expectancies that goes beyond the immediate concepts of interest, at least enough to differentiate them from the related local conceptual domain.

The task essentially is a conceptualization problem. In order to accomplish it we need theory. The key issue concerns where we are to find theories that enable rich pattern specifications that are reasonable. Theory can come from many sources: published literatures of relevant social or natural sciences; rational ruminations of the individual expert; examination of data or the observation of phenomena; or intuitions of program administrators, clients, and others who are party to an evaluation.

A central theme of this article is that no single source of theory is likely to be sufficient as the basis for articulating the patterns needed. Although "traditional" social science theory may inform pattern specification for program evaluation, it will almost never by itself be sufficient. This sentiment is similar to the point made by Chen and Rossi (1984: 339):

> It is an acknowledged embarrassment to our viewpoint that social science theory is not well enough developed that appropriate theoretical frameworks and schema are ordinarily easily available "off the shelf." But the absence of fully developed theory should not prevent one from using the best of what is already at hand.

Of course, the social sciences do not lack theories. The problem is that the theories they have are often not complete or specific enough for program evaluation. As Dunn (1981: 124) points out:

> There is no certain way to deduce causes from effects, or effects from causes, and social science theories are frequently so general or abstract as to be of little direct help in specific situations. In order to identify the possible causes contributing to a problematic situation, it is useful to have conceptualization frameworks that outline the many causes that may be operating in a given situation.

Chen and Rossi (1984: 339) also suggest what type of theory is needed:

> Of course, the kind of theory we have in mind is not the global conceptual schemes of the grand theorists, but much more prosaic theories that are concerned with how human organizations work and how social problems are generated. It advances evaluation practice very little to adopt one or another of current global theories in attacking, say, the problem of juvenile delinquency, but it does help a great deal to understand the authority structure in schools and the mechanisms of peer group influence and parental discipline in designing and evaluating a program that is supposed to reduce disciplinary problems in schools.

What is needed then are "smaller" microtheories that address issues more concretely within the contexts that are relevant. How are we to develop these more local microtheories? A seemingly obvious approach would be to incorporate more input from persons who experience the problems and run the programs on a daily basis and who thus could be expected to use at least implicit microtheories to guide their actions—administrators, program planners, clients, program staff, and other relevant constituencies. Chen and Rossi (1984:339) largely discount the value of this strategy:

> Nor do we argue for uncritically using the theories that may underlie policymakers' and program designers' views of how programs should work. Often enough policymakers and program designers are not social scientists and their theories (if any) are likely to be simply the current folklore of the upper-middle-brow media.

The contention here is that this position overstates the primacy of social science and seriously underestimates the value of attempting to discover the implicit theories that relevant constituency groups hold. Program administrators, clients, and staff probably utilize far more complex cognitive structures and micro-theories concerning the relevant social problems and programs than most social scientists do. It is not the absence of sensible microtheory but rather the difficulties of identifying and articulating it that is the central problem. How will we know if we have identified sensible microtheories? Chen and Rossi (1984:339) argue:

> The primary criterion for identifying theory in the sense used in this article is consistency with social science knowledge and theory. Indeed theoretical structures constructed out of social science concerns may directly contradict what may be the working assumptions of policymakers and program designers.

A very different position is argued here. The identification of theory must be based on *both* the consistency of the theory with social science knowledge and with the experiences and conjectures of constituencies that are directly familiar with the phenomenon in question. When contradictions do arise, primacy should not be given to established social science knowledge. Rather, the contradictory theory should be put to the test just as any new social science theory would be.

The management of the local theory development task can be viewed as a methodological problem. Methods are needed that go beyond traditional rational-analytic procedures. Evaluators will have to be capable of facilitating this conceptualization of micro-theory development task. This will necessitate the incorporation and balancing of conceptual input from a wide range of sources including the social science community, program administrators and staff, clients, and other constituencies. Although virtually every discipline has to address this issue of conceptualization, and many have developed techniques or strategies for doing so, there has been no coherent attempt to assemble these into a taxonomy of conceptualization methods that would be useful for evaluators. The next section describes initial steps in that direction and, for one methodology, illustrates how it might be utilized to enhance pattern matching program evaluation.

CONCEPTUALIZATION METHODS

A conceptualization method is defined here (somewhat broadly) as a specifiable or operationally defined procedure that, if followed, will result in a representation of what an individual or group is thinking. The result, the representation, can take many forms: it can be verbal, such as a speech or literature review on some topic; it can be mathematical, as in the statement of some formula or system of equations; it can be graphic, as in the case of a concept map or factor plot; or, it can be any combinations of these. Not all conceptual representations have utility for research purposes, especially for articulating the conceptual pattern expectations for an evaluation. We need to develop tradition about which methods work best for what types of situations.

At the outset, it is worth noting that there is no such thing as a "true" or "perfect" conceptual representation. Nor is it necessarily the case that we should think of a representation as reflecting some "true" conceptual structure with error. Cook and Campbell (1979: 63) echo a sentiment that is consonant with these thoughts:

> Our advocacy of multiple operationalism overlaps, therefore, with an advocacy of multiple formal definitionalism, provided that all definitions seem reasonable to most members of a given language group even though not necessarily accepted by all members of that group.

We are looking for useful conceptual representations that seem reasonable or can be interpreted, rather than for a "correct" one. In fact, for any evaluation, there may be many representations that would meet these criteria and be useful for pattern matching.

To get a better sense of what is meant by a "conceptualization method" we can, somewhat in passing, point to some of the better candidates (for pattern matching purposes) from several disciplines. The literature on such methods is diffuse and often the methods are not described as conceptualization methods per se. For instance, public policy analysts have a number of procedures that can be used to help "structure" policy problems. Some of these—classificational analysis, hierarchy analysis, synectics (Gordon, 1961; DeBono, 1973), brainstorming, and assumptional analysis (Mitroff et al., 1979) —are considered in an excellent review by Dunn (1981). Educational theorists have been developing freehand drawing of concept maps and the use of an epistemologically grounded "knowledge-V" (Novak and Gowin, 1984). A similar concept mapping strategy, termed "clustering," has been suggested as a method for conceptualizing essays and other writing (Rico, 1983). Evaluators and planners have utilized a wide range of methods under the heading "Delphi technique" (Helmer, 1966; Linstone and Turoff, 1975) to help achieve conceptual consensus in groups. Psychologists, systems theorists, and mathematicians have developed cognitive mapping strategies (Harary et al., 1965; Warfield, 1976; Axelrod, 1976) to articulate and examine causal relationships between concepts. Recently, several evaluators (Light and Pillemer, 1984; Cooper, 1984) have looked at how one might develop conceptual frameworks that can form the basis for conducting meta-evaluations of prior research literature. Content analysts have developed computerized approaches (e.g., key words in context) for extracting central themes or constructs from text databases, such as field notes or case records (Stone et al., 1966; Krippendorf, 1980). There is also a long and rich tradition in measurement and scaling that uses factor analysis (Rummel, 1970); multidimensional scaling (Kruskal and Wish, 1978; Davison, 1983); cluster analysis (Everitt, 1980; Anderberg, 1973); and Q-technique (Block, 1961) among others, to scale subjective judgments and often portray them in graphic form. Facet theory (Borg, 1979; Elizur and Guttman, 1976) is another methodology related to measurement theory that relies on articulation of "mapping sentences" that describe constructs and their interrelationships. The mapping sentence is related directly to a visual representation of the constructs obtained from multidimensional scaling analysis (Lingoes et al., 1977). Finally, recent efforts by the Joint Committee on Conceptual and Terminological Analysis of both the American Political Science Association and the American Sociological Association have led to development of a rule-based logical-analytical approach to elucidating important social science concepts (Sartori, 1984). All of these are conceptualization methods that could potentially be used to articulate the conceptual pattern expectations for a program evaluation, and together they illustrate well the diversity of literature and approaches that should be considered. Some of these methods are designed for individual use, some for groups, and some for either. Some require a specific form of input or a specific starting point; others are more open. Some yield results in text form whereas others generate graphs or numbers.

AN ILLUSTRATIVE CONCEPTUALIZATION

To illustrate what is meant by a conceptualization method and how it might be useful for the development of local or micro theory, a recent conceptualization study is presented. The study involved almost the entire staff of a University Health Service (UHS) organization in the development of a "concept map" that can be used for the generation of theoretical patterns in subsequent program evaluations. The conceptualization method relied on a group process that enabled the staff to state their views about their organization, and used multidimensional scaling and cluster analysis to generate a graphic representation of the group's thinking. As implied above, this approach is not the only method that could be used, nor is it necessarily the best. It is presented here to illustrate that evaluators can utilize methods that help to facilitate the articulation of implicit theories program developers and staff hold and that might form the basis of pattern-matching studies.

Procedure

The procedure followed here is described in greater detail in Trochim and Linton (forthcoming) and requires three major steps to accomplish the conceptualization, in this case the concept mapping. First, the participants generate a large set of statements that, as a group, describe the subject being conceptualized. Second, the participants provide information about the conceptual relationships between the statements. Finally, multidimensional scaling and cluster analysis procedures are applied to the relationship information to produce the "concept map" of the statements. This map is interpreted and can be used to help generate theoretical patterns useful for program evaluation.

Participants. A major goal of this process was to involve as many of the approximately 100 staff members of the UHS as possible, including maintenance staff, physicians, clerical staff, nursing staff, and administrators. In the statement-generation phase all staff were invited, and a total of 77 persons participated in separate sessions on three consecutive days. In the sorting phase, which was accomplished in two sessions, all staff were once again invited and a total of 69 staff members took part. The final meeting involved the interpretation of the concept map and included approximately 45 UHS staff members.

Generation of statements. The first step in this conceptualization involved having the participants brainstorm a large set of statements that, taken as a whole, describe how they view their organization. This was accomplished on three successive days with approximately 25 persons attending each session. On each day participants were divided into four or five small groups and were asked to

> generate statements (phrases or sentences) that describe your view of what University Health Services *should be* or *should do*. There are no correct, good, or bad ideas; all statements are good and valuable.

It is worth noting that this instruction is focused on the organization as a whole rather than on one specific program. It may be desirable in some studies to limit the conceptual domain more to reflect a specific program that will be evaluated. One person in each small group was designated the group scribe and was asked to record all of the items that were offered. Large sheets and marking pens were provided so that all group members would be able to see the items that were being recorded. A total of 315 statements were generated across the three sessions. Because of the relatively large number of statements, it was necessary to reduce the set to a number that was more manageable in subsequent steps. A final set of 100 statements was randomly selected from this larger set. The UHS administrator in charge of the project then examined the final set to see whether it appeared to be representative of the original set. No changes were made as a result of this examination. Each statement was then typed on an index card with an identification number and duplicate sets of the cards were made for use in the next step.

Sorting the statements. This step in the process was accomplished using an unstructured sorting procedure (Rosenberg and Kim, 1975). Each person was asked to sort the statement cards into piles "in a way that makes sense to you" with the only restrictions being that there could be neither N piles (i.e., 100) nor one pile. After this sorting task, each person was asked to assign numbers arbitrarily to the piles and to record each pile's number and the identification numbers of the statements that were sorted into that pile on a sheet that was provided.

For each person, a 100 × 100 binary symmetric matrix, X, was constructed from the sort of information where values in the matrix represent joint occurrences such that

$$X(i,j) = \begin{cases} 1 \text{ if statements i and j were sorted into the same pile;} \\ 0 \text{ otherwise.} \end{cases}$$

For instance, if a person put the statements numbered 7 and 35 in the same pile, a value of 1 would be entered in both row 7, column 35 and row 35, column 7 of the matrix. A 100 × 100 totals matrix, T(i,j), was then obtained by summing across the individual matrices. Thus any cell in this matrix could take on integer values between 0 and 69 (there were 69 participants who sorted statements) where the value indicates the number of people who placed the i,j pair in the same pile. High values in this total matrix indicate pairs of statements that were consistently placed in the same piles, whatever these piles might have meant to the sorters. The assumption here is that high values imply greater conceptual similarity. This matrix is used as input for the analyses in the next step.

Development of the concept map. Two general algorithms were used to produce the representation, in this case, the concept map. First, a multidimensional scaling (MDS) analysis (Kruskal and Wish, 1978; Davison, 1983) was conducted. Kruskal and Wish (1978: 7) describe the purpose of such an analysis:

> Multidimensional scaling, then, refers to a class of techniques. These techniques use *proximities* among any kind of objects as input. A proximity is a number which indicates how similar or how different two objects are, or are perceived to be, or any measure of this kind. The chief output is a spatial representation, consisting of a geometric *configuration of points*, as on a map. Each point in the configuration corresponds to one of the objects. This configuration reflects the "hidden structure" in the data, and often makes the data much easier to comprehend.

The specific analysis used was nonmetric multidimensional scaling, which requires only that the proximities be on an ordinal scale. A two-dimensional solution was selected. Second, a divisive hierarchical cluster analysis (Everitt, 1980; Anderberg, 1973) was conducted. This type of analysis begins by assuming that all statements are in a single cluster and then successively partitions the statements into smaller and smaller groups. The analysis is hierarchical, which means that once statements are divided into subgroups they can never be recombined again at later stages of the iterative process. The analysis utilizes a stopping rule to determine the number of clusters appropriate for the data (although theoretically the analysis could be carried out until each statement is a different cluster). Because the cluster analysis is hierarchical, the number of clusters selected is somewhat arbitrary and depends largely on the degree of cluster generality desired. The results of this hierarchical cluster analysis led to the selection of 12 clusters of statements. The original statements listed by cluster are shown in Table 1 for the first 3 of the 12 clusters.

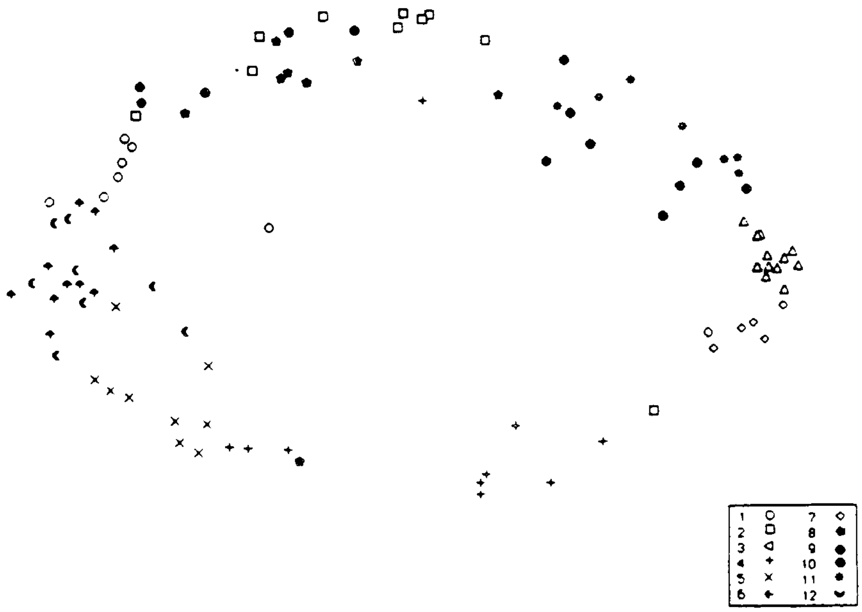
The results of these two analyses (MDS and cluster analysis) were combined to produce the final "concept map" shown in Figure 1. Each point on the map represents one of the 100 statements. The *location* of each point was determined by the x-y coordinates obtained from the multidimensional scaling analysis. The *symbol* that is used to plot the point shows the cluster that the statement is in as determined by the cluster analysis.

A second map was drawn by hand from the map in Figure 1 in order to show the approximate locations of the clusters of statements. This map will be referred to as the "cluster map" so that it can be distinguished in the discussion from the concept map in Figure 1. The group next went through a discussion process (first small group and then whole group) that led to consensus on names for the clusters that were then written on the cluster map. Finally, the group considered whether there were any sensible regions (i.e., groups of clusters) on the map. They agreed upon and named four general areas. The cluster names, region boundaries, and region names are filled in on the cluster map given in Figure 2. The group felt strongly that one of the clusters (cluster 5: Health Education) belonged simultaneously in two regions—Campus-Community Relations and Client Services—as shown in Figure 2. The cluster map is particularly useful as an aid for interpreting the concept map.

TABLE 1
Original Statements for Three of Twelve Final Clusters

Cluster 1 – Effective Patient Flow
Increase appointment times for patients (longer appointments)
Have a fast, same-day low wait service
Prompt services
Extend service hours
Team approach to seeing patients. Assign patients to team alphabetically
Discourage nonemergency visits to clinic during nonappointment hours
More interdisciplinary involvement
More staff availability to answer client questions
Cluster 2 – Improved Systems and Data Management
Centralized dictation for charts
Better computer system: (a) more cross-training on computers, (b) more access to computers by staff
Centralized dictation available to everyone in building
Increase computerization
Updated equipment
Improve transfer of medical records–decrease lost charts, decrease misfiled lab slips
Implement new system to give lab results
Improve system for dispensing medications
Increase educational activities
Cluster 3 – Staff Morale
Find ways to increase staff morale, especially during burn-out times (e.g., end-of-year crunch)
More acknowledgement of staff accomplishments; more positive feedback
Positive stroking
UHS should be supportive to staff
Less pressure on staff
Fewer hassles from administration
Increase sense of responsibility toward one's job
More openness regarding salaries
Promote salary equity–new employee vs. old employee
More integration among all UHS staff
Better orientation for new staff
Better communications among UHS departments

The final maps were interesting on several counts. The dimensions or directions on the maps appear to be directly interpretable—the dimension that moves from north to south could be described as moving from more managerial or administrative issues to ones that are more educational or service related. For instance, cluster 1 (Effective



NOTE: Each point represents one of the 100 statements. The location of the point is determined by the x-y coordinate information obtained from multidimensional scaling. The symbol used to plot the point represents which of the 12 clusters the statement is in as determined by hierarchical cluster analysis.

Figure 1: UHS Concept Map

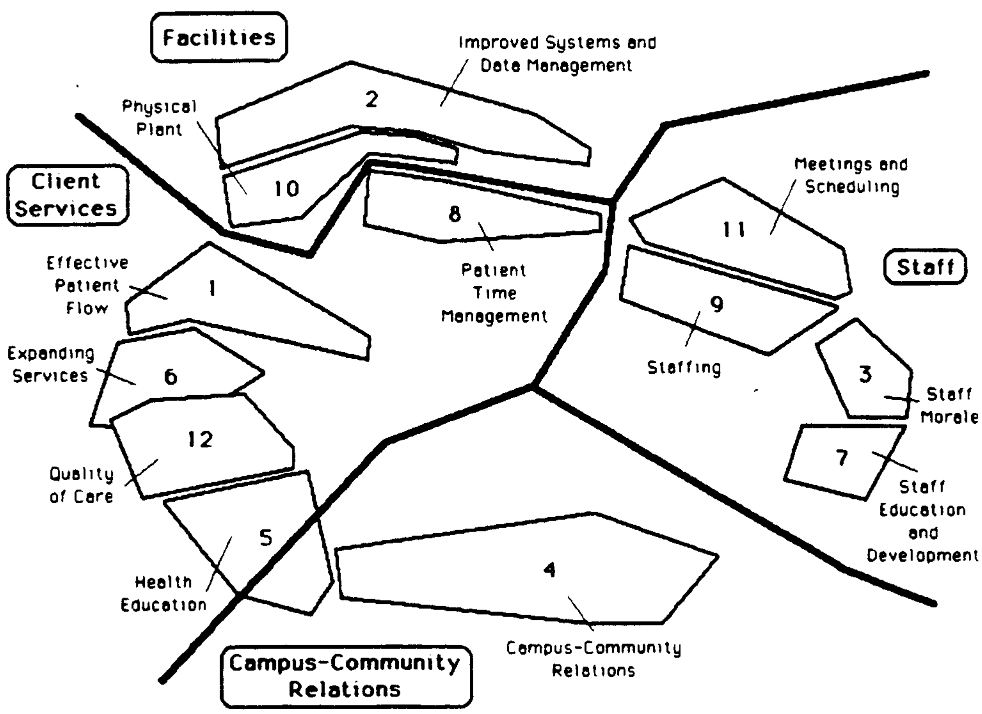


Figure 2: UHS Cluster Map Showing Approximate Locations of the 12 Clusters and 4 Regions Identified

Patient Flow) and cluster 11 (Meetings and Scheduling) lie to the north whereas cluster 5 (Health Education) and cluster 7 (Staff Education and Development) lie to the south. The east-west dimension is also easily interpretable with the western side representing more external, service, and client-related issues whereas the east reflects more the internal concerns of the staff. When viewed in terms of regions, the cluster map shows two areas that are related to the functioning of the UHS organization (facilities and staff) and two that refer to how the UHS relates to the outside environment (client services and campus-community relations).

The maps reflect the organization broadly, rather than emphasizing one particular program or service. This is largely because the initial brainstorming instructions were stated in terms of the general goals and mission of the organization. In any given program evaluation, it might be desirable to conduct a conceptualization that is more directly focused on the program in question. Nevertheless, because this organization wished to use the conceptualization as the basis for a general program planning and evaluation effort, a broad approach was chosen. As discussed below, these broad maps may have distinct advantages for pattern construction because they put any potential program within a larger conceptual context and anticipate some of the implications for the whole organization.

The maps can be considered the conceptual representation or theoretical framework within which one can construct programs, begin to develop strategies for construct validity, and articulate pattern predictions that can form the basis of program evaluation.

Using the concept and cluster maps to construct pattern matching program evaluations. We can best illustrate the potential utility of the maps through several hypothetical examples of program evaluation designs. First, let's consider that the organization intends to purchase a new computer system and wishes to evaluate the effects of this acquisition on the organization generally. Here, we can consider the computer system to be the independent variable, potential causal agent, or "program." We can begin to use the map to help guide the definition of the program (i.e., explore the "construct validity of the cause") by trying to locate the idea of the computer system somewhere on the cluster map. Clearly, the most appropriate location would be somewhere in the vicinity of cluster 2, Improved Systems and Data Management (we might even get more specific by looking item by item on the concept map in Figure 1). Examination of the statements in cluster 2 (see Table 1) suggests many of the issues of immediate relevance for the computer system implementation: the potential for the system to store and organize centralized dictation for case record charts; the need for training; the importance of access; issues related to equipment updating; and subsystems for managing medical records, lab information, and medication dispensation. Furthermore, the map indicates a broader conceptual context that can be considered and even suggests how related other conceptual issues may be (determined generally by distances between cluster 2 and the others). For instance, in developing the computer system, program planners should be aware of issues suggested in the most proximate clusters, in this case, cluster 10 (Physical Plant) and cluster 8 (Patient Time Management). This might lead them to greater awareness of the need to address space issues, placement of terminals in offices, or the development of a patient time management subsystem. The next closest clusters also suggest potential subsystems or areas of application: cluster 1 (Effective Patient Flow) and cluster 11 (Meetings and Scheduling). As we move further away from cluster 2, we see greater concern with overall quality of care, staffing issues, staff morale, and education of both staff and clients. We may assume that because they are more distant on the map, these areas will be less directly related to the implementation of the computer system itself. The most remote cluster, which would by definition be assumed to be at least relevant, is cluster 4, Campus-Community Relations. Note that the cluster map does not suggest that even this (or any other) cluster is irrelevant. Rather, it provides a network that suggests *degrees* of conceptual relevance. To assess whether the program (i.e. computer system and related activities) gets implemented as intended, the evaluator can construct measures or observations of the program activities undertaken in various cluster areas.

We can also use the maps to develop the theoretical measurement pattern. Within each cluster, the individual statements suggest potential operationalizations of outcome measures. For instance, the evaluator who wishes to measure cluster 1 (Effective Patient Flow) can get numerous measurement suggestions by perusing the statements in that cluster (see Table 1): numbers of appointment times and length of appointments; indicators of promptness of appointments (i.e., time lapse from presentation to appointment); available service hours of the organization; adequacy of assignment procedures (e.g., assignment to team versus individual); numbers of nonemergency visits during nonappointment hours; and so on. In this manner the evaluation could examine the degree to which implementation of the computer system is related to indicators of patient flow. Note that the map as a whole may be especially useful at suggesting "unintended" program effects. Without it, evaluators might not think to examine the effects of the computer system on constructs like "health education," "staff morale," or "campus-community relations," although effects in even these more remote clusters seem plausible. The distances between statements on the concept map may even suggest how strong the relationships between measures might be expected to be. Using these distances it might be feasible to construct a rough theoretical or expected correlation matrix for the relationships between measures.

The same concept map structure can also act as the framework for the development of a theoretical pattern of effects (i.e., the "construct validity of the effects") of introducing the computer system. We would, in general, expect to detect greater effects near the cluster most related to program implementation (cluster 2) with a declining pattern of effects as we move away from this. Thus, measures of change that reflect cluster 2 should show greater standardized differences than measures in, for instance, cluster 5, Health Education. More precisely, the distances between cluster 2 and the other clusters can suggest a hierarchical effect pattern (where smaller distances mean greater expected effects). Notice that cluster 2 is the central or focal cluster of relevance primarily for the physical computer system itself. If the program operationalization also includes a staff education effort, then we might consider the program focus to be located simultaneously in both cluster 2 *and* cluster 7, Staff Education and Development. In this case, patterns of effect would presumably emanate outward from both clusters. Assessment of the effects of both program dimensions (the computer system itself and the educational effort) would depend on the specification of these separate patterns. We would expect, for instance, that staff morale would be more affected by training efforts than by the acquisition of the physical system itself.

In developing these theoretical patterns, the concept and cluster maps are useful *suggestive* devices. They should not be treated as though they are definitive descriptions of implicit program theories or expectation patterns. Thus, as part of the interpretation process, the evaluator should explicitly encourage participants to challenge the locations of statements or clusters and the implied hierarchies of relationships. In fact, participants should be encouraged to move elements on the maps if such moves would lead to more satisfying or intelligible interpretations. The value of the maps stems from the rich conceptual or theoretical context they provide, not from their literal application.

The analysis of this hypothetical study would involve three pattern matches. First, the program match (program implementation) would be examined by looking at the relationship between the program as theorized and measures or observations of its implementation. One would examine what computer system elements (as identified at least in part on the map) were implemented and to what extent they were put in place. We would expect that the greatest implementation activity would occur on measures located closest to the focal cluster most germane to the program.

Second, the measurement pattern match (construct validity of the measures) could be examined using a variation of the multitrait-multimethod matrix approach. Presumably, measures would be taken at key locations throughout the map (with perhaps more measures collected near the focal cluster most related to the program). This match would examine whether correlations between measures that are closer on the map are higher than correlations that are farther apart.

Finally, the effect pattern match (causal hypothesis) could be examined by comparing the theoretical effect pattern (derived even in part from the examination of distances from the focal cluster as described above) and the obtained effects pattern. At a crude level, at least, we could assign greater expected effect values in the regions related to Facilities and Staffing than to those in the more distant clusters Health Education and Campus-Community Relations. At a finer level, we could actually rank the distances on the cluster map between the central program cluster and relevant effect locations and consider this ranking an inverse indicator of potential program effect (i.e., the farther from the central cluster some measure is, the less it is likely to be affected). The obtained effects pattern simply consists of the standardized measures of change that the evaluator constructs at different key points on the map. A strong relationship between these two patterns would suggest that the introduction of the computer system *caused* the pattern of effects to the extent that there are no clear competing explanations—methodological or substantive—that could adequately account for the effect pattern.

The statistical analysis would, of course, be complex. Multivariate analytic procedures are clearly called for, but even conducting numerous univariate statistical tests of change and ordering effect sizes by predicted causal pattern would help us to examine the relationship or correlation between the two patterns and may, at least roughly, allow us to confirm or deny a pattern match. As Campbell (1966: 101) points out:

> Actual statistics for estimating degree of pattern matching are, of course, not generally available, and the estimate of the human eye from graphed results is still the commonest criterion. The correlation coefficient sets a good example, however, by its equitable allocation of errors.

If a match is not apparent, we may need to revise (a) the "theory" or conceptual pattern as indicated by the maps; (b) our specification of the theoretical effect patterns (e.g., by reconsidering whether the program ought to produce that pattern of effects); or (c) our measures of effect. The strength of this design (and of all pattern matching designs) for program evaluation depends upon the degree to which the theoretical effects pattern is unique. If this pattern "matches" the obtained effects, there are likely to be few plausible alternative explanations for the obtained effect pattern and we can conclude that the program is the most likely causal agent.

Suppose, as a second example, that the organization felt it was important to address staff morale problems. In developing an appropriate program, they could begin by looking at items in the most relevant cluster, cluster 3 (Staff Morale). This would suggest a number of important ideas (see Table 1) related to positive feedback, supportiveness, salary issues, orientation sessions, and communications. Moving away from this focal cluster we find highly related issues suggested by the nearby cluster 7 (Staff Education and Development), cluster 9 (Staffing), and cluster 11 (Meetings and Scheduling). Thus we can use this conceptual map as the foundation for constructing both the theoretical program and the theoretical effect pattern. For the latter, we would again expect declining patterns of change as we move away from the central cluster related to staff morale.

From both examples we can draw several conclusions. First, the concept map provides a conceptual framework that can be used to construct the program theory, theoretical effect pattern, and expected relationships between measures. Second, the "distribution" of the program theory across the map may provide a hierarchical basis for the theoretical causal pattern. Third, measures of change that are collected across the map enable construction of the obtained effect pattern. Fourth, similarity between the theoretical and obtained effect patterns determines the degree to which construct validity of the cause-effect relationship can be claimed. Fifth, to the extent that no plausible alternative explanations for the obtained effect pattern exist, we can argue that the causal inference has internal validity.

The maps also force us to recognize other research problems that might not have been easily discernible. Most important is the recognition that a "program" may often be a complex conceptual pattern. Our traditional social research models that rely on dichotomous dummy-coded representatives of the program (or, at best, some indicators of actual implementation) are far less sensitive. Another problem that is clearly illustrated involves the confounding of programs or causal variables. If the organization only introduces a single change (e.g., a computer system), then it is reasonably possible to detect a pattern of effects. However, if at the same time they also include staff computer education and a program to improve staff morale generally, the combined expectation patterns become much more complex and the possibility of detecting such intermixed patterns becomes more remote. This shouldn't be viewed as a problem of the conceptualization itself, but rather as a real implication of investigating complex social programs.

These examples are not meant to imply that the scaling approach taken here is an exemplary method for accomplishing the conceptual task. On the contrary, there are many problems with this approach—methodological and substantive—that need further investigation and should lead us to be critical about its potential value for this purpose. For instance, it is reasonable to think that organizational theorizing alone might have led to the development of a conceptualization that is as good as or better than the maps presented here (in fact, we would hope that implicit organizational theories guided the participants at least in part in their generation of statements and in the sorting task). In addition, it is not even clear at this point whether a map is the best conceptual representation form for generating pattern predictions. The important points, however, are that (1) we need some well-conceived conceptual structure if we are going to conduct pattern matching studies; and (2) there are many conceptualization techniques or methods that we ought to investigate in order to determine their potential utility for pattern matching.

CONCLUSION

A few briefs statements can serve to summarize the case that is presented here. Pattern matching underlies all social research but for the most part the patterns typically used are rather crude and simplistic. More specific and detailed theoretical patterns will improve the construct validity of the cause (program implementation), the effect (measurement), and the cause-effect relationship. In addition, more complex patterns reduce the chances of finding plausible alternative explanations for an effect pattern, thus improving internal validity. Although traditional social science theory can help to inform program evaluation conceptualizations, it is likely that the evaluator will need to augment such theory using the insights and thinking of local program administrators and planners, clients, policymakers, and interest groups, among other sources of input. We need to develop a taxonomy of conceptualization methods that can help facilitate this task and we need to establish a tradition of testing the utility of such methods across many evaluation contexts. Finally, we need to develop statistical procedures that enable us to determine whether patterns are related.

The principles discussed in this article have implications extending beyond just causal hypothesis testing. Because the establishment of construct validity of the cause is directly related to program planning, the application of the thinking discussed here will be relevant for program and policy developers. Survey designs can also be improved. Closer attention to conceptualization and construct validity will enable the design of surveys where more complex response patterns can be anticipated through more detailed specification of the expected degree of item relatedness. Meta-evaluations can be improved considerably if they rest on a more detailed mapping of the expected pattern of relationships between studies. Finally, pattern matching can be used to enhance the credibility of other program evaluation designs. Both randomized experiments and other quasi-experiments besides the ones discussed here can be improved by incorporating richer theoretical patterns in place of more traditional dummy-coded program variables.

We are in a position to extend the logic of our research models through pattern-matching approaches. Whether we are able to exploit such advances depends on the degree to which detailed theoretical patterns can be specified in evaluation studies and on our ability to detect such patterns when they occur. Conceptualization methods can enhance the former; better statistical models for analyzing pattern-matching data are needed to improve the latter.

REFERENCES

ANDERBERG, M. R. (1973) Cluster Analysis For Applications. New York: Academic.

AXELROD, R. [Ed.] (1976) Structure of Decision: The Cognitive Maps of Political Elites. Princeton, NJ: Princeton Univ. Press.

BLOCK, J. (1961) The Q-Sort Method in Personality Assessment and Psychiatric Research. American Lecture Series 457. Springfield, IL: Charles C Thomas.

BORG, I. (1979) "Some basic concepts of facet theory," pp. 65-102 in J.C. Lingoes et al. (eds.) Geometric Representations of Relational Data: Readings in Multidimensional Scaling. Ann Arbor: Mathesis.

CAMPBELL, D. T. (1966) "Pattern matching as an essential in distal knowing," in K. R. Hammond (ed.) The Psychology of Egon Brunswik. New York: Holt, Rinehart & Winston.

———and D. W. FISKE, (1959) "Convergent and discriminant validation by the multitrait-multimethod matrix." Psych. Bull. 56: 81-105.

CAMPBELL, D. T. and J. C. STANLEY (1963) "Experimental and quasi-experimental designs for research on teaching," in N. L. Gage (ed.) Handbook of Research on Teaching, Chicago: Rand-McNally.

CHEN, H. and P. H. ROSSI (1984) "Evaluating with sense: the theory-driven approach," in R. F. Conner et al. (eds.) Evaluation Studies: Review Annual (vol. 9). Beverly Hills, CA: Sage.

COOK, T. D. and D. T. CAMPBELL (1979) Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

COOPER, H. M. (1984) The Integrative Research Review: A Systematic Approach. Applied Social Research Methods Series (vol. 2). Beverly Hills, CA: Sage.

CRONBACH, L. J. and P. E. MEEHL (1955) "Construct validity in psychological tests." Psych. Bull. 52: 281-302.

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

DEBONO, E. (1973) Lateral Thinking: Creativity Step by Step. New York: Harper & Row.

DUNN, W. (1981). Public Policy Analysis: An Introduction. Englewood Cliffs, NJ: Prentice-Hall.

ELIZUR, D. and L. GUTTMAN (1976) "The structure of attitudes toward work and technological change within an organization." Admin. Sci. Q. 21.

EVERITT, B. (1980) Cluster Analysis (2nd ed.). New York: Halsted.

GORDON, W.J.J. (1961) Synectics: The Development of Creative Capacity. New York: Harper.

HARARY, F., R. NORMAN, and D. CARTWRIGHT (1965) Structural Models: An Introduction to the Theory of Directed Graphics, New York: John Wiley.

HELMER, O. (1966) The Use of the Delphi Technique in Problems of Educational Innovations. Santa Monica, CA: Rand Corporation.

KRIPPENDORFF, K. (1980) Content Analysis: An Introduction to Its Methodology. Sage CommText Series (vol. 5). Beverly Hills, CA: Sage.

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

LIGHT, R. J. and D. B. PILLEMER (1984) Summing Up: The Science of Reviewing Research. Cambridge, MA: Harvard Univ. Press.

LINGOES, J. C., E. E. ROSKAM, and I. BORG [eds.] (1977) Geometric Respresentations of Relational Data: Readings in Multidimensional Scaling. Ann Arbor: Mathesis.

LINSTONE, H. A. and M. TUROFF (1975) The Delphi Method: Techniques and Applications. Reading, MA: Addison-Wesley.

MITROFF, I. I., J. R. EMSHOFF, and R. H. KILMANN (1979) "Assumptional analysis: a methodology for strategic problem solving," pp. 208-225 in L. Datta and R. Perloff (eds.) Improving Evaluations. Beverly Hills, CA: Sage.

NOVAK, J. D. and D. B. GOWIN (1984) Learning How to Learn. Cambridge: Cambridge Univ. Press.

RICO, G. L. (1983) Writing the Natural Way: Using Right-Brain Techniques to Release Your Expressive Powers. Los Angeles: J. P. Tarcher.

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

RUMMEL, R. (1970) Applied Factor Analysis. Evanston, IL: Northwestern Univ. Press.

SARTORI, G. (1984) Social Science Concepts: A Systematic Analysis. Beverly Hills, CA: Sage.

SHADISH, W. R., Jr. (1980). "Nonverbal interventions in clinical groups." J. of Consulting and Clinical Psychology 48: 164-168.

STONE, P. J., D. C. DUNPHY, M. S. SMITH, and D. M. OGILVIE (1966) The General Inquirer: A Computer Approach to Content Analysis. Cambridge: MIT Press.

TROCHIM, W. and R. LINTON (forthcoming) "Conceptualization for planning and evaluation." Evaluation and Program Planning.

WARFIELD, J. N. (1976) Societal Systems: Planning, Policy and Complexity. New York: John Wiley.

William M.K. Trochim is Assistant Professor in the Department of Human Service Studies in the College of Human Ecology at Cornell University. He is author of a book on quasi-experimental research design entitled Research Design for Program Evaluation: The Regression-Discontinuity Approach. *He has written on a wide range of topics related to program evaluation, including experimental and quasi-experimental research; statistical analyses for selection bias; program implementation; research quality control; statistical simulation; and conceptualization methods.*