PERGAMON

Evaluation and Program Planning 23 (2000) 211–230

EVALUATION and PROGRAM PLANNING

www.elsevier.com/locate/evalprogplan

Differences in stakeholder perceptions about training evaluation: a concept mapping/pattern matching investigation

Greg V. Michalski[a,*], J. Bradley Cousins[b]

[a]*Employee and Organizational Learning, Nortel Networks, Ottawa, Ontario, Canada K1N 4H7*
[b]*Faculty of Education, University of Ottawa, Ottawa, Ontario, Canada K1N 4H7*

Received 1 August 1998; received in revised form 1 May 1999; accepted 1 August 1999

Abstract

Concept mapping and pattern matching techniques were used in exploratory research to investigate differences in stakeholder perceptions of training results and evaluation in a major division of a multinational network-design and engineering company. Referencing a research framework informed by multiple constituency views of organizational effectiveness, this single case study examined group perceptions of the results and evaluation of training among 39 knowledge workers organized into three equal groups of 13. These groups consisted of (1) line managers as *sponsors* of training, (2) product developers as *participants* in training, and (3) training professionals as *providers* of training. A set of 100 statements were "brainstormed" by these groups to describe ideal results of training in terms of the success of the case organization. After sorting the 100 statements for conceptual similarity, all study participants also rated each statement's importance twice — once as a training result, and once as a training evaluation criterion. Using the concept maps developed by the groups, a set of 21 pattern matches were performed to investigate stakeholder differences both within (intra) and between (inter) groups. Correlation (Pearson r) coefficients were also calculated for each pattern match and displayed in tabular form for comparison. The overall results revealed that all stakeholder groups agreed reasonably well about the importance of training results in the organization. However, substantial differences were found regarding how each stakeholder group rated these results in terms of their importance for training evaluation. These differences were further found to correspond quite well with the unique organizational role of each stakeholder group. The results are discussed in terms of the potentials and limitations of concept mapping and pattern matching in training evaluation research, and their implications for training evaluation practice.

Keywords: Training evaluation research; Concept mapping and pattern matching; Stakeholder diversity; Power and politics; Organizational effectiveness

1. Introduction

Intellectual capital is defined as the sum of everything everybody in a company knows that gives it a competitive edge (Stewart, 1999). While training obviously represents a specific means to develop such capital, little has been done up to the present time to systematically evaluate the results of corporate-wide training programs in the era of intellectual capital (Conference Board Inc., 1997). In particular, training evaluation has been limited by an incomplete understanding of perceptual and expectation differences between key organizational stakeholder groups (McLinden & Trochim, 1998). This limitation is most acute in complex organizations staffed predominantly by knowledge workers[1] (Drucker, 1993a,b, 1995) who — as highly skilled and educated professionals with specialized roles and expertise — learn in many ways

* Corresponding author. Tel.: +1-319-341-2425; fax: +1-319-339-3020.

E-mail address: michalsg@act.org (G.V. Michalski).

[1] Peter Drucker originally coined the term "knowledge worker" *circa* 1960 to refer to educated professionals with high expertise and specialized organizational roles (see also Sumanth, Omachonu & Beruvides, 1990).

0149-7189/00/$ - see front matter
PII: S0149-7189(00)00005-7

in addition to formalized training (see, e.g., Brown & Duguid, 1991).

Stakeholder-based evaluation is useful to conceptually frame the general program evaluation problem (Alkin, Hofstetter & Ai, 1998). But most training evaluation approaches continue to ignore, diminish, or make broad and untested assumptions about the implications of stakeholder differences in terms of the design, execution, and consequences of training evaluation (Lewis, 1996). Among the main consequences of evaluation is the utilization of results. Evaluation utilization is the subject of a considerable body of theoretical and empirical literature within the broader domain of inquiry known as knowledge utilization. According to Shulha and Cousins (1997), scholars continue to think of the utilization of research findings or program knowledge in instrumental, conceptual, and symbolic terms. The stakeholder approach represents an appreciation that each program affects, in different ways, groups which have divergent and even incompatible concerns by realizing and legitimizing the diversity of interests at play (Weiss, 1983).

Although multiple stakeholder groups consisting of managers, trainees, and trainers have been recognized within the domain of instructional design (Broad & Newstrom, 1992), published (empirical) training evaluation studies involving these same three groups remain rare. One such study featured two of the three stakeholder groups mentioned above. In that study, Brown (1994) applied attribution theory to explore how managers and training professionals attribute causality for organizational results. He found that training professionals often identified training as either the sole cause, or a primary cause, of the results that had been achieved while managers rarely singled out training as a cause of such results. From a training evaluation perspective, Brown's innovative research is unique in focusing on the differing perspectives of managers — who can be viewed as the "paying" customers of training (Brinkerhoff, 1989) — and training professionals as program suppliers or providers internal to the organization. Yet current training evaluation practice continues to be largely limited to course-level trainee satisfaction, and to a much lesser degree, learning, and behavior changes (Bassi, Benson & Cheney, 1996).

Such mainstream training evaluation practice can be invariably linked to a taxonomy developed originally by Kirkpatrick (1975, 1994, 1996). This framework defines four levels of training results involving trainee (1) reaction, (2) learning, and (3) behavior, as the first three levels. The fourth level of the taxonomy focuses on the results of training at the business or organizational level. It has spawned numerous attempts to describe "level-four" training results specifically in terms of return-on-investment (ROI) and other financial measures (see, e.g., Fitz-enz, 1988; Geber, 1995; Kearsley, 1982; Phillips, 1991, 1996; Tesoro, 1998). While the assumptions of the four level taxonomy have been criticized (Alliger & Janak, 1989), and the extremely problematic task of attempting to quantify training results in financial terms continues to be debated (Abernathy, 1999; McLinden, 1995), discussions of training evaluation alternatives that engage the collective intellectual capital within knowledge-based organizations remain relatively scarce.

The current study used concept mapping and pattern matching (Trochim, 1989a) as an exploratory research technique to investigate stakeholder differences based on group perceptions of the results and evaluation of an entire training program as a collectivity of multiple, planned, learning interventions. The study was also guided by the research question: *To what extent do training stakeholder groups differ in their perceptions about the importance of training results and training evaluation*? The study was also guided by a research framework developed from the literature reviewed next.

2. Literature review

Training and its evaluation occur within a contextual setting. In the case of a large firm, this setting is the organization itself. While many attempts have been made to evaluate the results of training in financial terms, the complexity of organizations, and the myriad variables involved in their overall performance, often frustrate such attempts especially by overburdening them with assumptions (stated and unstated) that may or may not be testable. A central problem involves the extreme difficulty of isolating the effects of training and linking such effects to effectiveness measures. Yet the construct of organizational effectiveness has many potential measures best conceptualized in terms of a multiple-constituency perspective. Such a perspective fits well within a stakeholder-based evaluation approach. As a basis for the research framework developed to guide the present study, the following review focuses on two main domains of inquiry: (1) multiple-constituency perspectives of organizational effectiveness, and (2) stakeholder-based evaluation.

2.1. Multiple-constituency perspectives of organizational effectiveness

The theoretical precursors of organization effectiveness emerged in the early part of this century with the development of classical organizational theory (Spray, 1976). Since the establishment of these early roots, the literature has continued to grow substantially to include a broad and diverse range of theory and research.

Diverging from conventional and econometric indicators of organizational performance, Likert (1977) distinguished "traditional" theory from "modified" theory. The former is based on scientific management and cost accounting concepts. The latter is concerned with the human organization and concepts such as confidence, trust, motivation, loyalty and communication.

Organizational effectiveness has also been defined as a construct in terms of "goal-centered" and "natural systems" views (Campbell, 1976). The goal-centered view describes principal power centers or decision makers to define rational goals by objective measures and deductive means. The natural systems view abandons the possibility of meaningfully defining any small set of goals and embraces a more inductive mode of analysis. Categorizing several specific models having either a predominantly goal-centered or natural systems orientation, Campbell suggested combining the approaches to use the complementary insights provided by each. By reviewing empirical work on criterion measures of organizational effectiveness, Campbell listed a total of 30 such indicators. Among these are "goal consensus" defined as the degree to which all individuals perceive the same goals for the organization, and "training and development emphasis" defined as the amount of effort the organization devotes to developing its human resources (p. 37). Campbell stated that "In the best of all possible worlds, it would be nice to have some overall hierarchical map of how the criteria fit together in terms of their generality/specificity and means/ends relationships" (p. 39), and recommended continued research to describe the relationship between formal and operative goals using "several groups of individuals who might offer differing perspectives or expertise" (p. 41). The framework, design, and methodology used in the present study embraces this notion by seeking to reveal and explore such differing perspectives.

The notion of involving stakeholder groups engaged in brainstorming to generate statements to assess organizational effectiveness is not particularly new. In describing a multiple-constituency approach to assessing organizational effectiveness, Connolly, Conlon and Deutsch (1980) treat effectiveness not as a single statement, but as a set consisting of several (or many) statements, each reflecting the evaluative criteria applied by the various constituencies involved. Building on this work, Altschuld and Zheng (1995) reviewed several models for the evaluation of organizational effectiveness specifically in research organizations. These authors revisited Thompson's (1967) "Goal/Technology Contingency Table" and the use of social reference groups for effectiveness assessment. They stated "Lacking absolute criteria and causality related to outcome, complex organizations should turn to social referents to demonstrate their effectiveness" (p. 203). Essentially the satisfaction of constituent groups or individuals are indicators of an organization's effectiveness. The "competing values" framework for assessing effectiveness was presented to accommodate (in terms of open-systems theory) a range of organizational complexities including issues of values and politics. The competing values framework argues that there cannot be a single, universally acceptable model of organizational effectiveness. Concepts of effectiveness are value judgments based on an evaluator's personal beliefs, interests, and experiences. In assessing effectiveness, evaluators must determine what an organization's processes, behaviors, and ultimate goals are, or should be. In Campbell's words, "there is no algorithm of science that will specify which variables should be labeled as criteria of organizational effectiveness. That begins as a value judgment and ends as a political decision" (p. 40). Social knowledge construction (Bandura, 1986) provides an epistemological basis for multiple-constituency definitions of effectiveness. As an idea that has permeated numerous theoretical views on evaluation practice, stakeholder-based evaluation provides a means to derive and compare such definitions (Alkin, Hofstetter & Ai, 1998).

2.2. Stakeholder-based evaluation

Multiple-constituency and competing values approaches to effectiveness assessment support broadening participation and involvement in program evaluation. Stakeholder-based evaluation is an approach that identifies, and is informed by, particular individuals or groups. Stakeholders are the distinct groups interested in the results of an evaluation, either because they are directly affected by (or involved in) program activities, or because they must make a decision about the program or about a similar program (Gold, 1983; Guba & Lincoln, 1989; Stake, 1983). Alkin (1991) distinguished four different stakeholder roles. These include stakeholders as (1) primary users of evaluation results, (2) information sources for framing the evaluation, (3) data sources during the evaluation, and (4) the audience for the evaluation report.

In an effort to increase the responsiveness and utility for decision making, Gill (1989) proposed the adoption of a user-focused mode of training evaluation. Rather than have training evaluation remain the responsibility of training professionals, Gill advocated increased management involvement (see also, Nadler, 1984) to broaden participation. While expanded participation can promote utilization (Cousins & Earl, 1992), Patton (1997) stated that the use of evaluation will occur in direct proportion to its power-enhancing capability to "reduce the uncertainty of action for specific stakeholders" (p. 348). Patton also reviewed the technique

of stakeholder mapping, in which stakeholders can be categorized using a matrix according to their initial inclination toward the program (support, opposition, or neutrality) and how much they have at stake in the evaluation's outcome (high, moderate, or little), as a means to conceptualize stakeholder relations with respect to the program being evaluated. Program administrators, funders, clients, staff, and others are included among potential members of an "evaluation task force" assembled specifically to identify diverse perspectives. According to Patton, these members should represent "the various groups and constituencies that have an interest and stake in the evaluation findings and their use, including the interests of program participants" (p. 354).

The dimensions of stakeholder power and legitimacy were also recognized in stakeholder-based evaluation by Mark and Shotland (1985). They referred to power as the ability of a stakeholder to influence policy decisions involving the program being evaluated (e.g., program funders are high in power). Legitimacy refers to the reasonable interests of a stakeholder in one or more aspects of the program and its results (e.g., program participants). Although discussions of the influence of power are virtually absent in the training evaluation literature, this topic is much more well-developed in program evaluation practice and research (see, e.g., Chelimsky, 1987; Palumbo, 1987).

In summary, training evaluation practitioners have long sought to isolate benefits of training in terms of organizational effectiveness and business results. While many such efforts have been concerned exclusively with demonstrating the benefits of training from a singular stakeholder perspective in strictly financial terms, indicators of organizational effectiveness include a much broader range of possibilities involving multiple-constituency perspectives and stakeholder-based evaluation. These elements are combined and integrated to form the research framework developed for the present study.

The research framework is diagramed in Fig. 1. Main elements include the organizational (case) context, in which stakeholder groups are situated in terms of their views of training program results and evaluation. The organizational context includes multiple-constituency views of effectiveness from the perspective of knowledge worker employees possessing substantial intellectual capital within the firm. Three specific stakeholder groups are identified as training program (1) sponsors (line managers), (2) participants (trainees), and (3) training professionals. Both sponsors and participants are further depicted as training program clients, whereas, training professionals are shown as program providers. Each group is shown to view both training results and training evaluation from a distinctive stakeholder perspective. Individual group members are full-time employees who routinely interact with each other in relation to an established training program (specific details about the case organization, stakeholder groups, and training program are provided below).

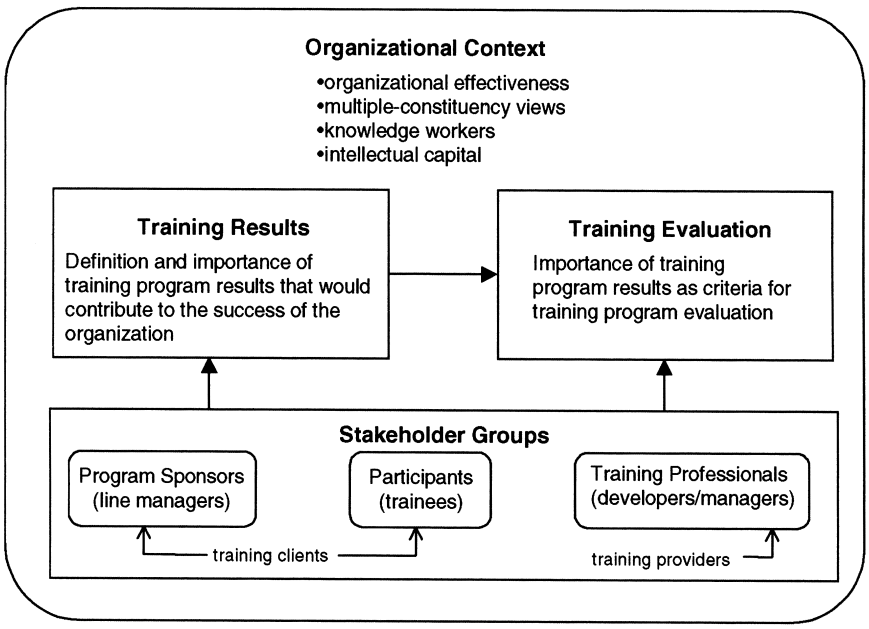

Fig. 1. Research framework.

3. Methods

Concept mapping was employed to define a domain of training results based on a set of brainstormed statements generated by three stakeholder groups. The use and discussion of this research technique in evaluation practice has been growing steadily over the past 15 years (Rizzo-Michelin, 1998) and, although various approaches to concept mapping have been developed and discussed in the literature, the approach described by Trochim (1989b) was selected for use in the present study based on its well-documented use in evaluation and program planning as well as its suitability for comparing group views using pattern matching techniques. Several authors have also described the use of this technique in both training evaluation and organizational studies (Cousins & MacDonald, 1998; Kolb & Shepherd, 1997). Before describing the details related to the sample, instruments, and the analyses used, the following section describes the case organization in which the present study was conducted.

3.1. Case organization

The study was carried out within a large division of a network-design and engineering company located in eastern Ontario, Canada. This company is co-located with a number of other technology-intensive firms, in a region known as *Silicon Valley North*. At the time of the study the division was staffed by approximately 1000 local, full-time personnel. Serving in professional job capacities as engineers, research scientists, software and hardware designers, and other highly educated and skilled (managerial, administrative, and technical) support staff, these employees are quintessential knowledge workers in the high-technology sector.

As part of a larger parent organization, a broad range of training, education, and development opportunities are available to these employees. Training within this particular division was formalized as a specialized, complete, and autonomous program designed to meet the specific needs of the divisional population. Specific instructional offerings consisted of a broad range of learning interventions and delivery modes including instructor-led, self-paced, computer-based, video, and audio instruction. Course offerings were organized into several specialized "training tracks" corresponding to major job responsibility areas. These areas include, for example, software, hardware, verification, and management development. At the time of the study there was no formalized evaluation process for the training program as a whole. Like most internal corporate training, individual course offerings were evaluated simply for participant satisfaction using familiar end-of-course feedback forms.

3.2. Sample

Based on an analysis of 38 concept mapping studies, Trochim (1993) found a range of 6–33 participants and noted the typically recommended sample size for concept mapping projects to be 15 people. For the purposes of the current study, three distinct stakeholder groups were defined in the case organization as training program (1) sponsors, (2) participants, and (3) providers. Because the total relative populations of individuals within each of these groups differed widely, a mixture of random, purposive, and self-selection was used to obtain the final sample of 39 participants in the study. From the population of approximately 900 trainees, 30 were randomly selected and invited to participate in the study. A total of 13 individuals agreed to participate. From the population of approximately 100 managers, 20 individuals known to be familiar with the training program were purposively selected and invited to participate. A total of 13 individuals agreed to participate. From the population of 25 training professionals[2] all were invited to participate. Again, a total of 13 individuals agreed to participate. Thus, the final sample consisted of a total of 39 program stakeholders arranged into three equal groups of 13.

3.3. Instruments and procedures

The preliminary concept mapping activity required members of each group to generate statements in response to a focus instruction regarding training program results. Each stakeholder group initially generated their own set of statements. Statements generated by each group were later combined into a single statement set using the procedure described below. The participant/trainee and training provider groups were involved in live brainstorming sessions facilitated by the principle author. Because of live group-scheduling difficulties, the program sponsor group (managers) generated brainstorm statements in writing (by E-mail or interoffice memo).

A single focus statement was used by all groups. This was a concise instruction directed to the group to stimulate brainstorming: *Generate statements (short phrases or sentences) that describe specific training program results that would contribute to the success of [the organization] over the next 12 to 24 months*. A focus prompt was used by respondents as a check while generating statements to stay on task. For the focus state-

[2] The principal author was a training professional in the case organization at the time of the study. Bias risk was minimized because he was not among the 13 training providers in the sample.

ment (above) each brainstormed statement should make sense when preceded by the following focus prompt: *One specific training program result that would contribute to the success of [the organization] over the next 12 to 24 months is....*

Following typical brainstorming protocol (open and free flow of ideas, no censoring, say what you think, etc.) a relatively large number of raw statements was initially obtained. This included a total of 219 unedited statements (72 sponsor, 77 trainee, 70 provider). These were entered into a single database and tagged according to which group generated the statement. This raw statement set was edited down to a final set of 100 statements. The editing procedure involved a combination of processes to (1) remove many obvious redundancies (e.g., many statements were made that described positive training results in terms of employee and customer satisfaction, quality, project execution, business sector competitiveness, profitability, etc.); (2) clarify and render consistent terminology (e.g., acronyms and abbreviations); and (3) correct spelling and grammar. Throughout this editing and reduction process the proportionality of the original set of raw statements was preserved so that approximately one-third of the 100 statements came from each of the three groups (as in the raw set).

All sorting and rating was done by each study participant as an individual activity via internal mail by distributing a complete package of required materials. This package contained (1) a two-page set of instructions, (2) one set of 100 cards (each containing a statement from the final set) for sorting, (3) a sort recording form, (4) two statement importance rating forms, and (5) a self-addressed return envelope. Each data form was coded to identify the respondent by group. The precise instructions for sorting and rating statements followed those typically used for the concept mapping technique employed (see, e.g., Concept Systems Inc., 1996a,b; Trochim, 1989b). These instructions directed each study participant to perform two main tasks.

The first task was to sort the statements into individual piles in a way that made sense to the individual (and to record the results of this sorting on the sort recording form provided). Thus, sorting was done by organizing each of the 100 numbered statement cards into piles of similar meaning from the perspective of the sorter. Having done this, the sorter was further instructed to record all sorting results by indicating the statement numbers contained in each of his or her final piles. Each of these statement piles was then labeled by the sorter with a short phrase or title describing the contents of the pile (e.g., a pile containing a dozen statements related to project execution might be labeled as "project management" by the sorter). All responses were recorded on the sort recording form to complete the sorting task.

The second task involved rating the importance of each statement twice: once as a training *result* and once as a training *evaluation criterion*. This task was performed by having each respondent rate the importance of each of the 100 statements on two separate forms. The following instruction was used:

> *Rate the importance of each statement on two separate forms. Each form contains exactly the same set of statements and rating scale. The only difference is that on one form you will rate the importance of the item as a training result, and on the other form you will rate the importance of the item as a training evaluation criterion.*

All statements were rated using the following five-point rating scale: 1 = Relatively Unimportant; 2 = Somewhat Important; 3 = Moderately Important; 4 = Very Important; 5 = Extremely Important. The analysis procedures used are described next.

3.4. Analysis

All sorting and rating data provided by each of the 39 respondents were analyzed as a single project using Concept System software.[3] This software provided a convenient means to perform the statistical calculations used to generate the initial concept maps, to represent and refine these maps based on stakeholder group input, and to derive the subsequent pattern matches. The major calculations performed by the software include sort data aggregation, multidimensional scaling, cluster analysis, bridging analysis, and sort pile label analysis. Detailed discussion of these calculations is beyond the scope of the present paper. However, both the main statistical procedures — multidimensional scaling (MDS) and cluster analysis — and the application of such in concept mapping have been well-described elsewhere (see, e.g., Anderberg, 1973; Kruskal & Wish, 1978; Davison, 1983; Everitt, 1980; Trochim, 1989b).

Three individual stakeholder group concept maps were computed initially using a default number of six clusters per map. Because respondents supplied individual pile labels along with their statement sorting results, it was possible to extract initial concept map cluster names using the top ten best pile label algorithm. The software derives these after the map is computed using the pile names supplied by the sorters. This is done by calculating the "best fitting" sort pile label for a cluster based on centroid

[3] Version 1.71 of the software was used for this analysis (Concept Systems, Inc., 1996b).

computations. Briefly stated, the best pile label algorithm uses MDS results to perform a distance calculation that compares the distance between each cluster centroid and each sort pile centroid. The best sort pile label is the one closest (in Cartesian distance) to the cluster centroid.

Both the six cluster (default) maps and the initial labels were used only as aids during concept map interpretation sessions. Three separate "live" concept map interpretation sessions were held to allow each group to discuss, interpret, and label the clusters on its map. The results of these interpretation sessions produced the final concept maps (discussed shortly). Along with the rating data, these maps also formed the basis for subsequent pattern matching analyses.

Pattern matching allows for the combination of any two measures (e.g., statement importance ratings either within or between groups) aggregated at the concept map cluster level to examine the degree to which the measures match. By examining such combinations of measures, differences between stakeholders can be identified and quantified. A pattern match always involves two patterns based on measurements taken at the statement level, for example, as importance ratings using one or more scales. Pattern matching is powerful in its implications, particularly as a measure of stakeholder consensus regarding their relative views of statement importance within specific concept map clusters.

The results of a pattern match are represented both graphically (as a ladder graph) and numerically (as a correlation coefficient) between measures. The ladder graph is comprised of two vertical scales — one for each measure — joined by variably sloping lines each corresponding to a labeled concept map cluster. If the match is perfect, all lines will be horizontal and the resulting graph resembles a type of ladder (the lines connecting the clusters have zero slope). Such ladder graphs are especially useful for quickly spotting "disconnects" between two measures. The correlation coefficient associated with each pattern match can range between -1 and $+1$. It is a standard Pearson r (product–moment correlation) between the average ratings of the two variables, and it is useful to describe the strength of the relationship between them. Pattern matching is typically performed after concept mapping using statement rating data.

Both the theoretical basis and practical application of pattern matching have been well described (McLinden & Trochim, 1998; Trochim, 1985, 1989d, 1990). Pattern matching can be used to examine the degree of consensus between two groups by comparing the statement ratings of one group with those of another for any concept map cluster. For example, if a particular stakeholder group were to produce a concept map with five clusters, with each cluster containing 20 statements rated on two different scales, pattern matches could be derived to compare the average importance ratings of statements contained in each cluster in two distinctive ways. This could be done either (1) within the group by comparing importance ratings for the same statements in each cluster between the scales, or (2) between groups by comparing importance ratings for the same statements in each cluster between the groups. For the current study, pattern matching was done both within (intra-group) and between (inter-group) stakeholder groups using the three concept maps and two statement importance rating measures. A total of 21 such pattern matches were performed.[4] Three intra-group pattern matches were derived to compare each group's ratings using the two statement importance rating scales. Another 18 pattern matches were developed to make inter-group comparisons. These are discussed next.

4. Results

The training results conceptualized by each stakeholder group are depicted in three corresponding concept maps. Each map shows the 100 statements as numbered stackings. Numbers correspond to statement numbers. The height of each stack corresponds to the average training result importance (TRI) assigned to the statement by the group. Each numbered stacking is organized within the concept map as cluster groupings. Clusters were labeled by each group during three separate concept map interpretation sessions.

In addition to being rated in terms of TRI, the same statements were also rated in terms of training evaluation importance (TEI). This resulted in another set of three concept maps with the same cluster configurations, but with point and cluster stackings corresponding to TEI (rather than TRI). To respect page limitations all concept maps rated in terms of TEI are not shown, but these are compared to the TRI maps presented using the pattern matching correlations (presented shortly). The set of 100 statements, along with the average importance ratings assigned by each stakeholder group, appears in the table in Appendix A. Details about the stress and bridging values computed for each of the three concept maps appear in Appendix B.

The concept map results produced by each stakeholder group are presented next. These results are in-

[4] While the theory and application of pattern matching has been documented (as discussed) the authors are not aware of other published studies in which multiple pattern matches have been used to compare more than two groups as in the present study.

itially compared within each stakeholder group by three intra-group pattern matches using data from the two rating scales (TRI-TEI). A series of 18 inter-group pattern matches were developed to examine differences between the stakeholder groups. These are discussed following the individual group results.

4.1. Training providers

Training providers perceived relatively fewer, but conceptually broader, categories compared with the other two stakeholder groups. The training provider concept map, showing statement and cluster importance ratings in terms of TRI, is shown in Fig. 2.

Five conceptual clusters are shown. The rank order of clusters in terms of TRI is (1) benefits resulting from training (M = 3.72), (2) customer value (M = 3.71), (3) effective training program attributes (M = 3.38), (4) employee satisfaction (M = 3.34), and (5) skills and knowledge (M = 3.29). These results emphasize the benefits and necessity of training in terms of achieving the goals of the organization, particularly in traditional terms of productivity and efficiency. Recognizing that all statements describe *specific training program results that would contribute to the success of [the organization]* it is notable that the training provider group specifically labeled clusters as "benefits resulting from training" and "effective training program attributes" on their map. Among the statements ranked highest in average TRI were those supporting an ostensibly managerial or business perspective. This included (♯3) support customer's strategic and operational objectives (M = 4.85), (♯73) makes new people productive as quickly as possible (M = 4.38), (♯36) better support of strategic direction of organization (M = 4.31), and (♯11) training program objectives obtained directly from organizational business objectives (M = 4.08). In contrast eight of the ten statements ranked lowest in TRI by this group were more closely aligned with trainee, rather than management, concerns. Examples include (♯42) increase relevant training days per staff (M = 2.00), (♯38) employee professional/educational credentials are better recognized (M = 2.31), (♯34) employees say "there's so many good courses and so little time" instead of "are there any courses I should take?" (M = 2.62), (♯72) employees look forward to training to learn something new and "neat" (M = 2.69).

In sharp contrast to their statement ratings as training results (TRI), training providers rated TEI much differently. In rank order of TEI the average cluster ratings were (1) effective training program attributes (M = 3.63), (2) skills and knowledge (M = 3.46), (3) employee satisfaction (M = 3.34), (4) benefits resulting from training (M = 3.25), and (5) customer value (M = 2.95). These results stand in stark contrast to the importance of the same statements as training results.

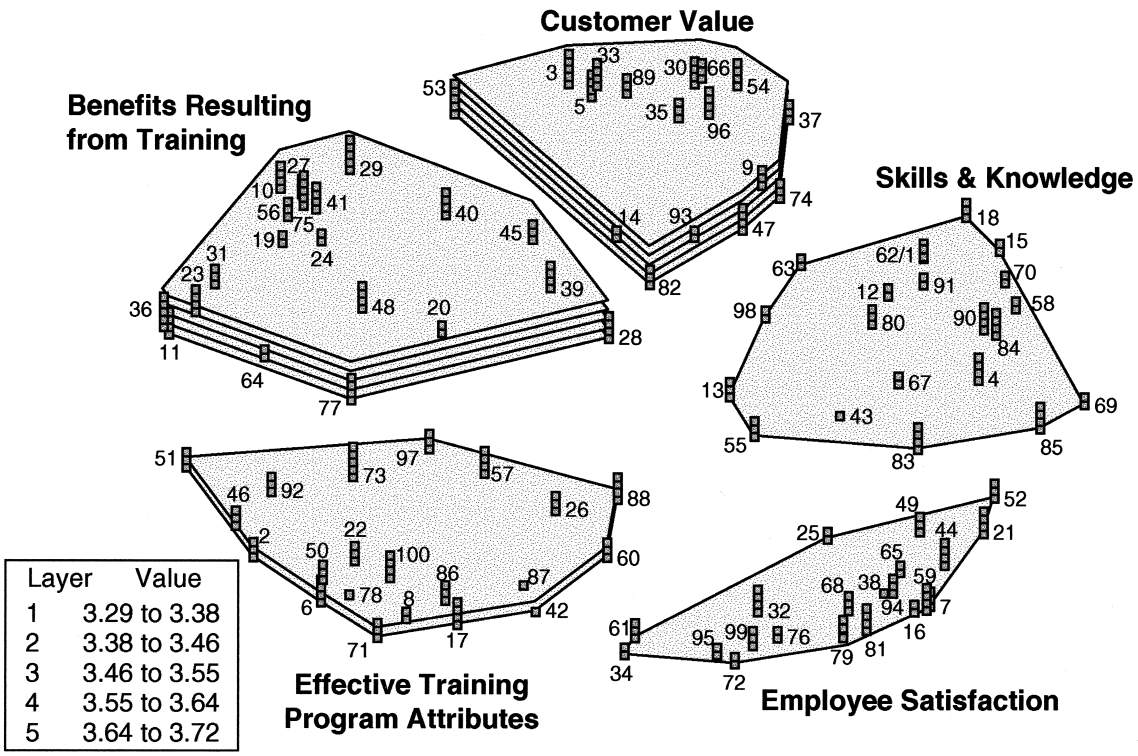

Fig. 2. Training provider group concept map rated in terms of TRI (*Note*: Numbers 1–100 correspond to statements shown in Appendix A).

This contrast is depicted graphically and quantified as the intra-group pattern match shown in Fig. 3.

This figure also serves to represent the relationship between concept mapping and pattern matching by showing the concept map produced by the provider group as rated in terms of TRI (left map) and TEI (right map). As shown, cluster statements rated in terms of TRI contrast markedly with this group's ratings of the same statements in terms of TEI.

As noted, the pattern matching correlation coefficient is a standard Pearson r correlation between the average ratings of two variables, in this case, the TRI scale on the left and the TEI scale on the left. A negative cluster-level correlation ($r = -0.75$) between the TRI and TEI ratings suggests that while the training provider group tends to view training as an important ingredient in benefiting the organization — especially in terms of contributing to customer value and satisfaction — they do not deem these same results to be the most important ones in terms of training evaluation. Rather, as the ladder diagram shows, this group views the TRI of the cluster statements inversely compared with ratings of the same statements in terms of TEI. Included in the "effective training program attributes" cluster are those statements strictly referenced to training, teaching, learning, and the integration and improvement of such. By considering this conceptual cluster to be highest in TEI, the training provider group seems to view training evaluation as somewhat parallel to the evaluation of their own performance in delivering these results as program providers. This result is examined further in the discussion section.

4.2. Program sponsors

Training program sponsors conceptualized training results especially in terms of market competitiveness, customer satisfaction, and product development. Fig. 4

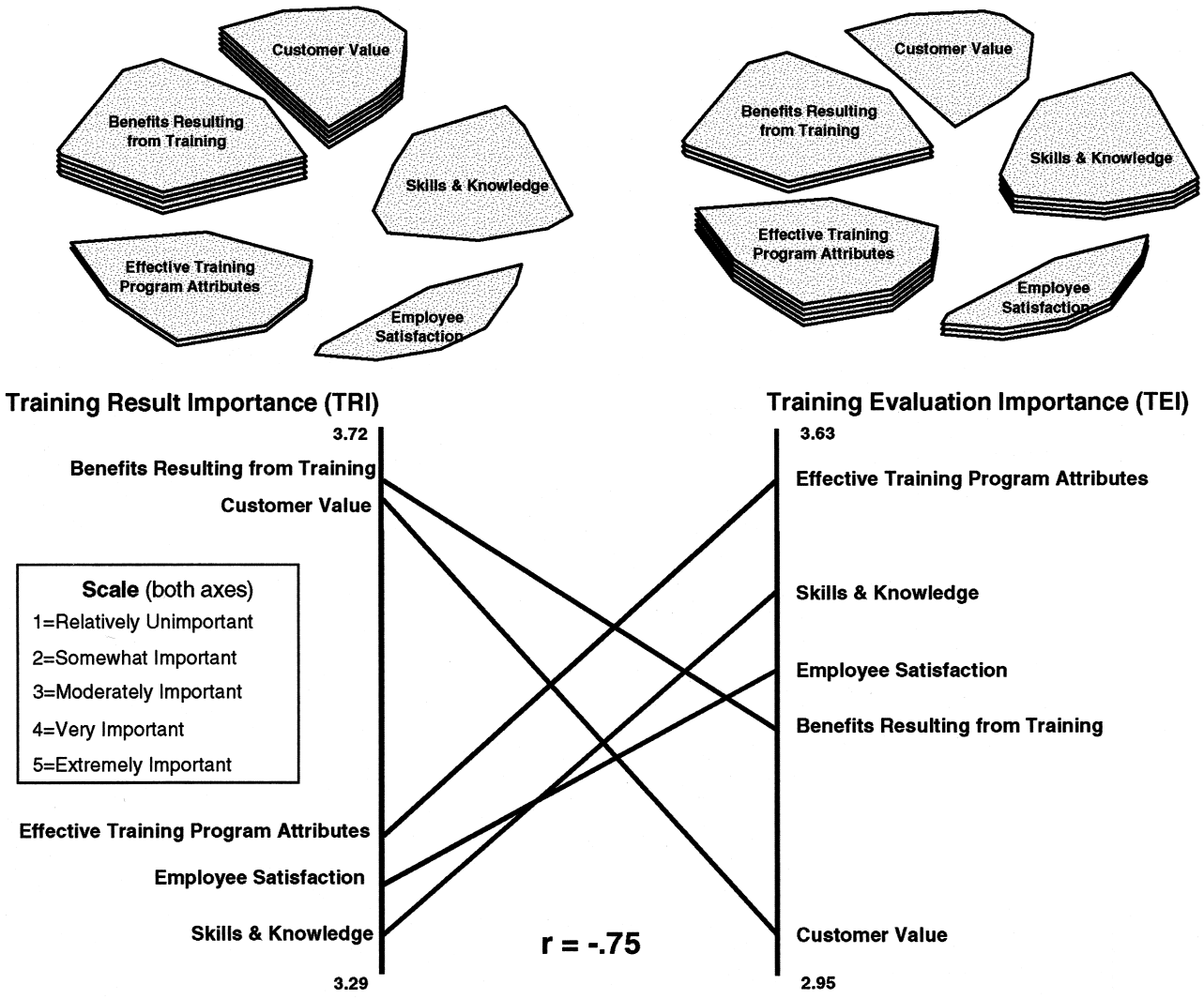

Fig. 3. Training provider intra-group pattern match illustrating relationship between concept mapping and pattern matching.

shows the concept map (rated in terms of TRI) produced by the program sponsor group.

The map shows six clusters. As ranked in order of average TRI these clusters included (1) customer and market (M = 4.22), (2) product development (M = 3.71), (3) collaboration and knowledge alignment (M = 3.59), (4) employee development (M = 3.50), (5) organizational/corporate (M = 3.46), and (6) training-learning integration (M = 3.02). The top two conceptual clusters emphasize this group's close transactional proximity to the interface between the external customer and internal product development and staffing priorities. Among the statements ranked highest in TRI by this group were (♯75) reduced time to market (M = 4.77), (♯30) ability to meet/anticipate customer requirements (M = 4.62), (♯96) improve ability to turn product capabilities into value for the customer (M = 4.46), and (♯81) attracts/retains key employees (best & brightest); develops employee loyalty (M = 4.46). In contrast, the statements ranked lowest by this group included those which do not directly or immediately relate training to product development or customer satisfaction. Examples, include (♯8) training is integrated with university–industry interaction programs (M = 1.92), (♯6) individual learning style(s) are addressed by optimization of learning media (M = 2.31), (♯42) increase relevant training days per staff (M = 2.38), (♯94) employees get credit/acknowledgment/rewards for their learning achievements (M = 2.62), and (♯61) employees have increased control of their training and knowledge resources (M = 2.77).

Program sponsors tended to rate statements for both TRI and TEI in a much more concordant manner compared to training providers. Statements comprising the "customer and market" cluster were rated highest in both average TRI (M = 4.22) and TEI (M = 3.46). This similarity is summarized by a fairly strong, positive, intra-group pattern matching correlation coefficient (r = 0.74). A particularly notable difference is that, rather than ranking "product development" next in TEI, this group ranked "employee development" (M = 3.35) second overall. These results are discussed further below.

4.3. Participants/trainees

The participant/trainee concept map is shown in Fig. 5.

This group produced the most highly defined map which included nine conceptual clusters. In rank order of TRI, these clusters were labeled as (1) design quality (M = 3.88), (2) customer and market orientation (M = 3.68), (3) project preparedness (M = 3.40), (4) employee satisfaction (M = 3.39), (5) people management (M = 3.38), (6) process awareness (M = 3.35), (7) organizational training support (M = 3.28), (8) business management (M = 3.15), and (9) learning improvement (M = 3.14). This group emphasized the project-driven job perspective of the employee as

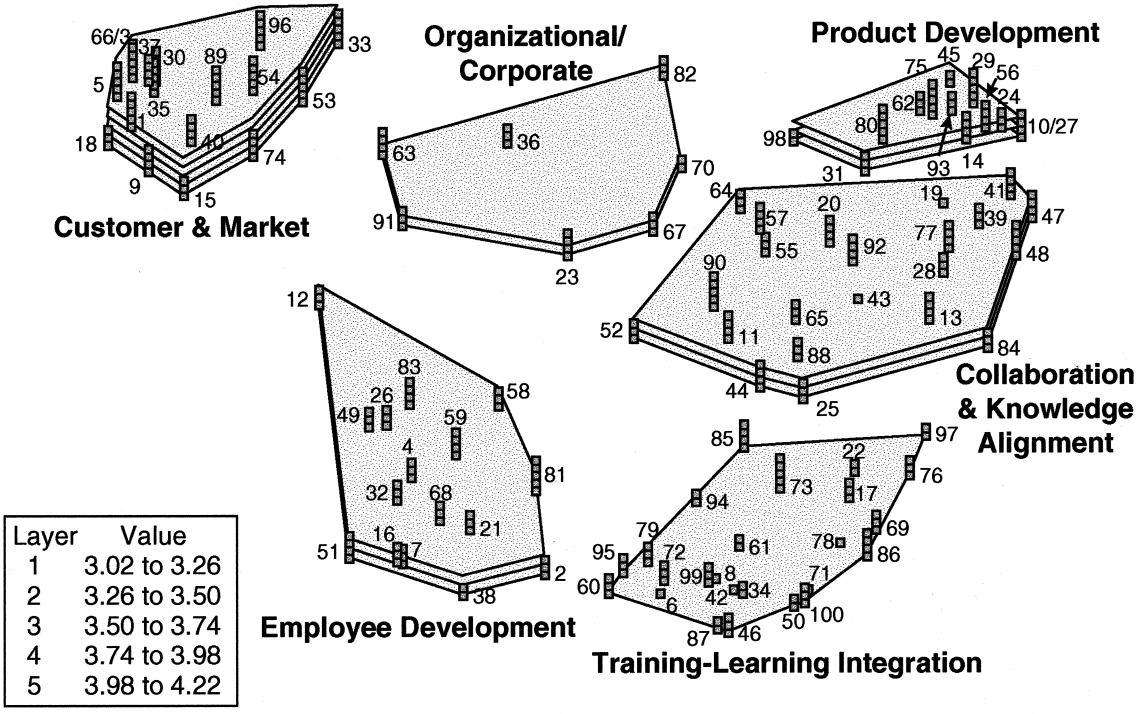

Fig. 4. Training sponsor group concept map rated in terms of TRI (*Note*: Numbers 1–100 correspond to statements shown in Appendix A).

designer/developer, particularly in terms of design, customer, and quality processes. Both of the top two clusters contain statements that emphasize the essential role of this group's expertise as core knowledge workers in the organization. These results suggest that this group views itself as a main vehicle toward achieving many of the organizational results with training acting as a catalyst. Individual statements ranked highest in TRI were (♯81) attracts/retains key employees (best & brightest); develops employee loyalty (M = 4.46), (♯29) significant product quality improvements (M = 4.31), (♯48) less time correcting mistakes/fewer recurrent problems (M = 4.23), and (♯75) reduced time to market (speed deliverables, reduce design cycle times) (M = 4.23). Among the statements ranked lowest in TRI were those emphasizing more bureaucratic or managerial themes. These include (♯78) attain training program compliance with standards (M = 2.54), (♯43) contribute to line management perception of staff project preparedness (M = 2.54), (♯42) increase relevant training days per staff (M = 2.62), (♯51) faster transition to 90% effectiveness for new managers (M = 2.69), (♯69) obtain "certified" special skills (M = 2.77), and (♯23) support key performance plan of organization (M = 2.77).

The trainee group also rated the statements for TRI and TEI more concordantly than training providers. In parallel to this group's TRI ratings, statements in the "design quality" cluster were also ranked first in TEI (M = 3.64). "Project preparedness" (M = 3.42) and "process awareness" (M = 3.25) were rated next in TEI. An examination of the statements in these clusters would seem to indicate that this group values training program results as outcomes that contribute directly to (or support) their employee expertise. In contrast to the training provider group (but similar to the program sponsor group), the participant/trainee group rated TRI and TEI in a fairly analogous manner. The participant/trainee intra-group pattern match shows a fairly strong positive correlation (r = 0.65) between statement ratings for TRI and TEI. This result is notable because — in contrast to the training program provider group — both the participant/trainee and program sponsor groups (as training program clients) view the results and evaluation of training as being much more closely aligned. Both client groups showed much better agreement between how they conceptualized TRI and TEI relative to the training provider group (based on the intra-group pattern matching results presented). Thus, both client groups would seem to expect training to be also evaluated in terms of its ability to deliver the results they expect from it. As discussed earlier, the training provider group does not share this view. We now examine more closely comparisons across the three stakeholder groups.

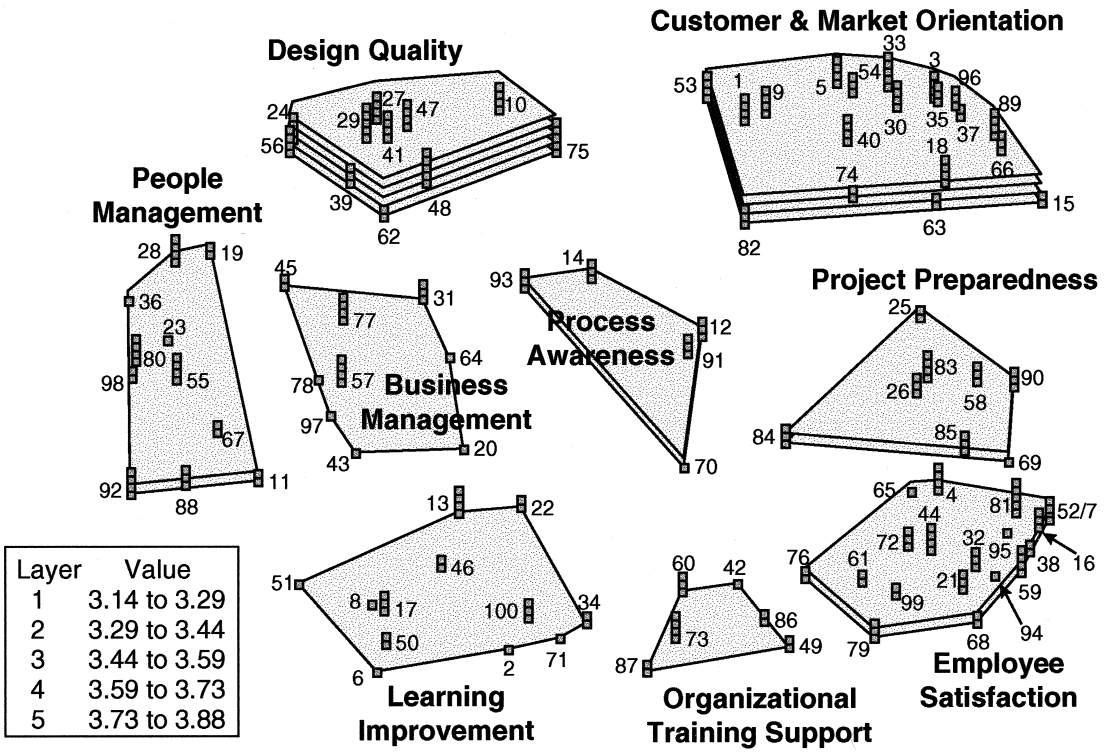

Fig. 5. Participant/trainee group concept map rated in terms of TRI (*Note*: Numbers 1–100 correspond to statements shown in Appendix A).

4.4. Inter-group pattern matching results

As discussed, in addition to its utility for describing differences within each stakeholder group (using ratings from the two scales), pattern matching techniques can also be used to explore differences between groups. In addition to the 3 intra-group pattern matches discussed, a total of 18 inter-group pattern matches were also performed to describe differences between the groups. These pattern matches compare the importance ratings of statements within the clusters of any stakeholder reference map to the importance ratings of the same statements as rated by the other two stakeholder groups. Rather than show these pattern matches graphically (by showing the concept maps and ladder diagrams together as in Fig. 3), a table of correlation coefficients is used as an efficient presentation means. A side-by-side examination of these tabular coefficients is useful to quickly reveal group differences. In addition to the three intra-group values discussed Table 1 presents all inter-group pattern match correlation coefficients.

Compared with the other two stakeholder groups, training providers generally demonstrated the strongest contrasting views regarding training evaluation. This table shows fairly good agreement concerning inter-group views of training result importance among all groups. Strong to very strong positive pattern match correlations ranging from $r = 0.73$ to $r = 0.90$ were obtained by comparing ratings for TRI to TRI across the groups. Moderately positive pattern match correlations (ranging from $r = 0.24$ to $r = 0.49$) were also obtained by comparing TRI to TEI ratings across the program sponsor and participant/trainee groups only. The strongest negative inter-group correlations occurred between the training providers and both other (training client) groups in pattern matches comparing TRI to TEI and TEI to TEI. For example, comparing the program sponsor TRI ratings to the TEI ratings of the same statements by the training provider group, a pattern match correlation of $r = -0.94$ was obtained. This same comparison (TRI-TEI) between the participant/trainee and training provider groups yielded a pattern match correlation of $r = -0.40$. An inter-group comparison between the program sponsor and training provider groups for TEI (TEI-TEI) yielded a pattern match correlation of $r = -0.57$. Only one mildly negative correlation ($r = -0.28$) was obtained comparing the ratings of program sponsors to participant trainees for TEI. These results then demonstrate and quantify differences in stakeholder views of training results and training evaluation. The implications of these findings are discussed next.

5. Discussion

The purpose of this study was to investigate differences in stakeholder perceptions about training evaluation using concept mapping and pattern matching as an exploratory research technique. As such several aspects of the study are discussed particularly in terms of its (1) research contributions and limitations, and (2) implications for training evaluation practice.

5.1. Research contributions and limitations

The study contributes to a relatively small but emerging body of stakeholder-oriented training evaluation research. As traditional approaches to training evaluation (e.g., those based on Kirkpatrick's four level taxonomy and its successors) continue to influence common practice, published research focusing on stakeholder diversity is a relatively recent — and still underrepresented — topic in the training and human performance literature. Training evaluation approaches and models based upon stated or unstated assumptions regarding training results from a singular (or limited) stakeholder perspective are increasingly being discussed

Table 1
Pattern matching correlation coefficient matrix

	Pattern match correlation coefficients (r)[a]								
	Training provider			Program sponsor			Participant/Trainee		
Stakeholder reference map	TRI-TRI	TRI-TEI	TEI-TEI	TRI-TRI	TRI-TEI	TEI-TEI	TRI-TRI	TRI-TEI	TEI-TEI
Training provider	–	−0.75[b]	–	0.78	−0.01	−0.37	0.81	0.48	−0.12
Program sponsor	0.87	−0.94	−0.57	–	0.74[b]	–	0.80	0.24	−0.28
Participant/Trainee	0.73	−0.40	0.17	0.90	0.49	0.58	–	0.65[b]	–

[a] *Note*. All correlation coefficients are derived from cluster-level pattern matches referencing each stakeholder map (column). Each correlation relates the average cluster importance of statements in the stakeholder reference map to the average importance of the same statements as rated by the other two stakeholder groups.

[b] Intra-group pattern match.

as vestiges of the industrial age (Pepitone, 1995; Stewart, 1999). Such approaches have become anachronistic in an age of intellectual capital because they explicitly or implicitly ignore or diminish the voices of knowledge workers in the evaluation process.

Yet the importance of stakeholder diversity has been recognized in more recent models of training evaluation. In describing his model for thinking about the evaluation of training, Lewis (1996) noted that such models reflect a point of view. According to Lewis, how we conceive of training and evaluation has to do with our view of the purposes of training. From what he defined as a *corporatist* (utilitarian) perspective, training evaluation will be held at the same level of discourse as buying a machine. Such a view obviously propagates efforts to focus on training evaluation in strict financial terms of ROI. Alternatively from the *humanist* perspective, the views of workers and their specific interests both within and beyond the workplace are considered key. By recognizing and balancing multiple-stakeholder perspectives in his training evaluation model, Lewis effectively bridges the gap between traditional training evaluation approaches (suitable to the command and control organizations of the past century) and the new realities of the intelligent organization.

The current study also contributes to the growing volume research employing concept mapping and pattern matching. Although concept mapping has been widely used in training evaluation and related studies (Cousins & MacDonald, 1998; Moad, 1995; Setze, 1997), published empirical studies expanding the use of concept mapping, and particularly pattern matching, remain scarce. The approach used in the current investigation should provide a useful reference point for comparison, especially using pattern matching techniques, across a range of cases and organizations. The expansion of such empirical work would provide a further benefit as a means to test and apply concept mapping and pattern matching theory (Trochim, 1989c).

Several limitations should also be noted. These apply both specifically to the study itself and more generally to the concept mapping techniques used. First, as noted above, few empirical studies of stakeholder differences in training evaluation are available for comparison. Furthermore, because the present investigation was performed as a single case study within a single division of a larger organization, the generalizability of the results is limited. Additional trials and applications of the techniques used here are both needed and encouraged. Second, although going beyond the limited practice of involving only one or two distinctive stakeholder groups, the current study was limited to only three such groups. Consequently the question can be legitimately raised as to expanded definition and involvement of other potential organizational groups and even sub-groups. Additional stakeholder groups could be identified, for example, as sales and marketing, manufacturing, and other distinctive professional organizational groups. All such groups could also be further subdivided by various other characteristics such as a specific cultural identity, gender, or age classifications. Third, rather than randomly selecting the members of each of the three stakeholder groups, a mixture of random, purposive, and self-selection was used. To encourage participation in the study, the program sponsor sample was derived from a subset of managers who were known to be familiar with the training program. Also, although they were randomly identified as potential study participants, in the interest of ethical research practices, members from all three groups were essentially self-selected. That is, all potential study participants randomly identified were under no obligation to participate and a number of those identified, in fact, did choose not to participate.

In addition to the limitations of the study itself, several general limitations of the concept mapping technique used should also be briefly mentioned. Of those issues raised elsewhere (see, e.g., Rizzo-Michelin, 1998), we resonate particularly with those listed by Kolb and Shepherd (1997). These involve four main areas related to general limitations involving the (1) reliability and stability of concept mapping results over time, (2) lack of a well-developed means for comparative research especially due to the relative dearth of available pattern matching studies, (3) relatively non-intuitive nature of the concept mapping process, which can lead to confusion about statement sorting and map interpretation for various participants, and (4) challenges of organizing and coordinating large-scale concept mapping projects in terms of the logistics of implementation. From a wider perspective none of these limitations should be construed as fatal flaws. Rather, the first two can be readily addressed through continuing research, and the last two can be addressed by improved facilitator knowledge and skill in the concept mapping process. In addition to the research contributions and limitations mentioned, the study also suggests at least two major interrelated implications for training evaluation practice. These involve the continued fixation on ROI measures and the role of organizational power and politics related to training. They are considered next.

5.2. Implications for training evaluation practice

The first implication calls into question the popular tendency to evaluate training using short-term financial measures of ROI. Contrary to such popular notions, little evidence was produced (even among managers)

to support any strong expectations to evaluate training for its ROI. There were several statements that referred to training results in terms of cost (e.g., statement ♯31-reduced cost of operation internally, ♯46-realize increasingly cost-effective training), but ROI was generally not emphasized.[5] Also each stakeholder group conceptualized these cost-related statements quite differently. For example, statement ♯31 (about internal cost reduction) was conceptualized as a benefit resulting from training by the training provider group. This same statement was conceptualized in terms of product development by the program sponsor group, and in terms of business management by the participant/trainee group. Rather than conceptualize cost reduction as the direct benefit or result of training, both of the two latter groups included the statement in clusters that characterized it more as an indirect or subsidiary result. Recognizing that generally accepted accounting principles do an unacceptable job of accounting for the principal activities of knowledge-intensive businesses (Stewart, 1999), these results probably reflect well upon the case organization because its members seem to have (wisely) resisted the popular call for training ROI.

The second implication suggests the need for an increased awareness, recognition, and even explicit mention of the role of organizational power and politics in training evaluation. As noted, the topic of organizational power and politics is notoriously absent from discussions in the training evaluation literature.[6] Perhaps it is exactly because such evaluations have largely ignored stakeholder-based approaches that such considerations are rare. Yet in making clear the pervasiveness of politics in evaluation, Weiss observed that stakeholder-based evaluation "...takes evaluation down from the pedestal and places it in the midst of the fray" (1983, p. 11) and that "The politics of program survival is an ancient and important art" (1987, p. 49). Yet beyond any evaluation purposes stated in objective terms, power and politics must play a role — even in training evaluation.

Some tentative evidence for the presence of power and politics was revealed by the present study. The group differences revealed seem to relate and correspond quite well with the organizational role of each stakeholder group. Pattern matching results comparing the training provider group with the other two groups are illustrative. Recognizing that training providers may perceive themselves to have the highest stake in the program (relative to the other groups), the pronounced pattern match differences suggest a strong alignment with the goals and values of their internal clients (i.e., sponsors and trainees). Training providers conceptualized TRI in a way that both emphasized their involvement in key organizational results, and highlighted the value of their services as support staff. For example, the statement ranked highest overall by this group in TRI was statement ♯3 — support customer's strategic and operational objectives ($M = 4.85$). In terms of organizational power dynamics, this result fits quite well with Mintzberg's (1983, p. 138) description of professional support staff in which he stated

> *...they work in small, fractionated groups offering rather vulnerable services to the organization (since these can usually be bought externally), it is in their interest not to pressure for autonomy but rather the reverse — to encourage their involvement in the decision processes. Collaboration is important to the professional support staff.*

But, while ostensibly seeking to be perceived as collaborating with other stakeholders to help bring about positive organizational results, the training provider group did not seem (based TRI-TEI pattern matching results) to favor being seen as exclusively accountable for delivering these results through training alone. This suggests that training providers might support a more evidence-based, and less cause-effect oriented, approach to training evaluation. This can probably be attributed to the group's heightened recognition of the complexities of "proving" training's contribution in a complex organizational setting as noted by McLinden (1995, p. 15):

> *Specifically, studies that are designed to focus solely on the extent to which interventions affect bottom-line indicators ignore the way organizations work. That is, a myriad of other effects can occur between the intervention and the measurement of an effect on fees, profits, customer satisfaction, and other indications. Disentangling the myriad variables, isolating and unequivocally proving the single effect due to training, may simply not be possible.*

Additional evidence of this can be seen by examining the statement ratings in the map clusters and pattern match shown in Fig. 3. The cluster labeled "customer value" is a good example. This cluster was ranked second in overall TRI (by only 0.01 units less than the top-ranked cluster), yet was ranked lowest overall in terms of TEI. An examination of the differences in the individual statement ratings (e.g., ♯3, ♯30,

[5] Although the study was not specifically designed to look for ROI support this result is, nonetheless, notable given the substantial number of references (noted) that describe training ROI as the primary concern of corporate management.

[6] There are, however, a few noteworthy exceptions including Darrah (1995), and Goleman (1998, p. 248) who briefly mentions "organizational politics" as a reason why "pet" training programs do not get evaluated in any substantive way.

♯33, ♯47, ♯53) in this cluster supports the apparent provider group belief that training contributes to — rather than causes — key organizational results. Because training providers seem to recognize that causal proof is effectively impossible (or perhaps too expensive) to obtain, the same cluster statements that were rated high (M = 3.71) in terms of TRI were rated much lower (M = 2.95) in terms of TEI. Hence, as professional support staff this group may recognize the political importance of aligning and collaborating with their internal clients, but they also seem to be ever-challenged to produce appropriate (low or no cost) indicators of training's contributions in doing so. This finding resonates well with Goleman's (1998, p. 249) comments:

> *Rather than an objective assessment of the effects of training, the typical evaluation comes in the form of "happy sheets," feedback from participants about whether trainees liked the program, and what about it they liked the most — a rating systems that patently favors slick, fun experiences over substantive ones. Having a good time becomes the mark of excellence, a valuing of entertainment over education.*

Turning now to the program sponsors, this group is key precisely because of position of formal authority in the organization. According to McLinden (p. 13) "The belief is that evaluation involving both those with substantial position power and those with close transactional proximity to the training is necessary to provide compelling evidence of strengths and weaknesses". As noted, the program sponsor group was very consistent in how they rated TRI and TEI based on their concept mapping/pattern matching results. By far this group considered training results in terms of customer and market to be most critical. The statement ranked highest in TEI was statement ♯73 — makes new people productive as quickly as possible (faster ramp-up, e.g., to 90% efficiency) (M = 4.08). This result suggests a strong concern with organizational change and/or growth. Assuming the need to bring in "new people" is not just to stem normal employee attrition, the perceived importance for training to contribute to the productivity of new employees portends a concern with organizational growth — by definition, "new" employees are those brought in and added to the population of existing employees. This corresponds well with organizational power theory. According to Mintzberg (1983) managers invoke power and political influence based on the system of authority characterized by position power, budgeting discretion, and accountability. The needs of the line managers are reflected in two different forces: on the one hand, an identification with the CEO and his goals of survival and growth, especially at higher levels in the hierarchy; and, on the other, an attempt to satisfy their own drives for autonomy and achievement through the aggrandizement of their own units and the balkanization of the overall structure. Both forces, however, favor expansion of the organization at large, and so growth emerges as the key goal among managers.

Participants/trainees comprise the central core of knowledge workers in the organization. This group researches, designs, prototypes, tests, develops, and maintains highly technical networking products and services that provide the main business revenues to the company. This group owns a large percentage of the organization's collective intellectual capital. Participants/trainees emphasized the importance of training results and evaluation mostly in terms of their own professional practice. This group rated concept map clusters labeled in terms of design quality and project preparedness as the most important. According to Mintzberg, the power and political basis of this group directly involves the system of expertise (see also Pfeffer, 1981, 1994; Toffler, 1990). This is characterized by a normative reward system and high intrinsic satisfaction. Group goals involve (1) protection and autonomy of the group, (2) enhancement of the prestige and resources of the specialty and professional excellence (sometimes in spite of client need), and when client–professional relationships are close and personal (3) support of the organization's mission. As engineering and design professionals, group members tend to take pleasure in their work and would be expected to place a higher relative value on TEI indicators related to their own career and professional growth and job satisfaction.

6. Conclusions

As a means to more fully utilize the collective intellectual capital of various knowledge worker groups in training evaluation, the expanded use of concept mapping and pattern matching seems promising. Yet it is important to recognize that further work needs to be done to adequately probe the stakeholder differences surfaced here. For example, given that (1) training professionals are routinely called upon to either lead or participate in evaluations of their own programs and that (2) other stakeholder groups can add value to this process, the results produced have obvious relevance within wider discussions about internal (Love, 1991) and participatory (Cousins & Earl, 1992, 1995) evaluation. But, even within these, conclusions based solely on the present study would be premature. Beyond the replication across a wider variety of cases mentioned, continuing research should further investigate stakeholder group views related specifically to the purposes,

Table A1
Average training result importance (TRI) and training evaluation importance (TEI) of statements by stakeholder group

		Average (M) statement importance by stakeholder group[b]					
		Training providers		Program sponsors		Participants	
	Statements[a]	TRI	TEI	TRI	TEI	TRI	TEI
1.	Increase understanding of current data networking needs and industry directions	3.38	3.08	4.46	3.62	4.08	3.69
2.	Simplify the learning efforts required of newly-promoted managers	3.31	3.23	3.54	3.62	2.85	2.31
3.	Support customer's strategic and operational objectives	4.85	3.15	4.38	3.46	4.08	3.00
4.	Better employee interpersonal and communication skills	3.77	3.38	3.46	3.31	4.08	3.46
5.	Products exceed customer requirements (e.g., features, operability, cost, maintenance ease, quality of service)	3.92	3.00	4.23	3.38	3.92	3.38
6.	Individual learning style(s) are addressed by optimization of learning media	3.77	4.23	2.31	2.69	2.77	3.00
7.	Contribute to employee sense of professional (self)marketability	3.46	3.62	3.15	3.00	3.54	3.54
8.	Training is integrated with university–industry interaction programs	2.85	3.23	1.92	1.77	2.85	3.08
9.	Promote designer understanding of big picture (how their product fits into market)	3.46	3.38	3.69	3.46	4.08	4.00
10.	Decrease maintenance effort/cost of completed software (e.g., decrease calls to helplines)	3.85	3.15	3.92	3.31	3.85	3.54
11.	Training program objectives obtained directly from organizational business objectives	4.08	3.85	3.69	3.54	3.15	3.15
12.	Develop corporate awareness	3.08	2.38	3.31	3.00	3.38	3.23
13.	Improve "effectivity" of teamwork (more effective staff)	3.69	3.69	4.08	3.31	3.77	3.31
14.	More accurate estimates	3.08	2.38	3.92	3.31	3.62	3.46
15.	Develop the equivalent skills that we would receive if we spent 1 year working for customer's business	3.08	3.38	3.62	3.62	3.23	3.00
16.	Increase employee confidence	3.08	2.85	3.15	2.92	3.38	2.77
17.	Just-in-time, just-the-right, training needs are addressed	4.23	4.08	3.62	3.85	3.54	4.00
18.	Better ability to understand the drivers and supporting technology of the telecommunication industry	3.62	3.38	3.62	3.62	3.69	3.31
19.	Reinforces linkage of product development processes to "company" specific processes	3.00	3.00	2.46	2.54	3.15	2.69
20.	More efficient organization with improved capacity to take on a broader range of activities	3.08	2.54	3.69	3.54	2.92	2.85
21.	Support personal broadening & job enrichment (develops employee interests in future tasks/roles/responsibilities)	3.92	3.54	3.62	3.69	3.46	3.23
22.	Iconic (isolated) training ceases and integrated training begins	3.15	3.54	2.62	2.92	3.15	3.15
23.	Support key performance plan of organization	3.92	4.23	4.08	3.77	2.77	2.54
24.	Improve modular software design methodologies (code structured for modularity)	2.92	2.62	3.54	3.46	3.85	3.92
25.	Prepares staff for development programs not yet launched (anticipatory education)	3.08	3.08	3.31	3.46	3.08	3.38
26.	Enables a more proactive (positive) response to change	3.62	3.31	3.62	3.15	3.38	2.85
27.	Decrease in number of defects per lines of executable code	3.54	3.31	4.00	3.38	3.77	3.85
28.	Sharing of product development knowledge (open exchange of ideas) is fostered	3.77	3.46	3.54	2.92	3.77	3.31
29.	Significant product quality improvements	4.31	3.38	4.46	3.77	4.31	3.77
30.	Ability to meet/anticipate customer requirements (even when the customer doesn't know them)	4.23	2.92	4.62	3.46	4.08	3.38
31.	Reduced cost of operation internally	3.15	2.92	3.46	2.92	3.54	3.00
32.	Stimulates high level of employee interest & motivation to continue learning	4.00	4.08	3.31	3.38	3.46	3.85
33.	Customer satisfaction improves at a higher rate	4.15	3.15	4.23	3.46	4.15	3.15
34.	Employees say "there's so many good courses and so little time" instead of "are there any courses I should take?"	2.62	2.77	2.77	3.62	3.23	3.77
35.	Better exposure to customers networks and business plans	3.31	2.31	4.31	3.23	3.38	2.69
36.	Better support of strategic direction of organization	4.31	3.85	3.62	3.62	2.85	2.69
37.	Build and run networks end to end to permit valued customer propositions	3.15	2.38	4.15	3.15	3.15	2.54
38.	Employee professional/educational credentials (existing expertise) are better recognized	2.31	3.08	2.77	2.46	3.23	2.62
39.	Better hand-off between functions (e.g., h/w designers to board layout; s/w designers to verification)	3.77	3.15	3.38	2.69	3.54	3.46
40.	Improve ability to leverage change in the technology to the best advantage of our company and our customers	4.15	3.46	4.15	3.38	4.00	3.62
41.	Improve "evolvability" of designs	3.77	2.85	4.00	3.31	4.08	4.00
42.	Increase relevant training days per staff	2.00	2.85	2.38	2.77	2.62	2.54
43.	Contribute to line management perception of staff project preparedness	2.38	2.38	2.46	2.15	2.54	2.15
44.	Personal productivity increases	3.92	3.23	3.69	3.46	3.69	3.31
45.	Develop awareness and understanding of new product development process (e.g. IPI — integrated product introduction)	3.69	3.69	2.85	3.00	3.31	2.62
46.	Realize increasingly cost-effective training	3.15	3.69	3.08	3.46	3.00	3.00
47.	Better ability to write efficient, high-quality, code	4.08	3.62	4.08	3.69	4.00	4.00
48.	Less time correcting mistakes/fewer recurrent problems (disseminates "lessons learned" knowledge)	4.23	3.38	4.23	3.85	4.23	3.23
49.	Better manager understanding of employee training and the MFA process	3.23	3.46	3.15	3.46	3.15	2.69
50.	Flexibility in sourcing world-class technical training	3.38	3.62	2.92	3.46	3.31	3.23
51.	Faster transition to 90% effectiveness for new managers	3.15	3.46	3.77	3.62	2.69	2.54

52.	Long-term goal setting skills of employees are improved	3.23	3.23	3.31	3.08	3.23	3.00
53.	Create "company" differentiator "Network Supplier of Choice"	4.38	2.62	4.62	3.08	3.77	2.69
54.	Narrow the gap between how a designer thinks and how our customers think	3.77	2.69	4.38	3.62	3.62	3.31
55.	Foster networking (between employees across departments and divisions)	3.62	3.54	3.62	3.00	3.85	3.23
56.	Allow changes to be made to existing software more easily	3.46	2.77	3.77	3.54	3.85	3.46
57.	Increase organizational learning (captures/redeploys intellectual property)	4.08	3.85	4.00	3.38	3.77	3.15
58.	Increase general telecommunications and computer (h/w, s/w) "literacy" of employees	3.08	3.69	3.23	3.46	3.62	3.85
59.	Employee satisfaction improved	3.77	2.92	4.08	3.38	3.85	2.69
60.	Learners become teachers and mentors to others	3.69	3.62	3.08	3.38	3.38	2.92
61.	Employees have increased control of their training and knowledge resources	3.38	3.15	2.77	2.92	3.15	3.08
62.	Better general understanding of the benefits/application of root cause analysis	2.77	2.85	3.23	3.08	3.00	2.69
63.	People appreciate business reality, not just their technical specialty	2.77	2.85	4.00	3.92	3.31	2.62
64.	Support our (organizational) approach to work and job design (structuring work for effectiveness)	2.92	2.69	3.15	3.00	2.85	2.54
65.	Facilitates employee job transitions between functions (e.g., s/w, h/w, etc.)	3.08	3.31	3.15	3.08	2.85	3.08
66.	People appreciate customer's viewpoint, not just their technical specialty	3.54	2.85	4.46	3.85	3.46	2.92
67.	Improve employee understanding of organizational "culture" (how they fit-into organization)	2.92	2.69	3.23	3.31	3.00	3.31
68.	Demonstrates that the company is investing in the employee and his/her career	3.69	3.77	3.38	3.54	3.23	3.00
69.	Obtain "certified" special skills (e.g., project mgt. or code inspection)	2.85	4.15	3.15	3.38	2.77	3.08
70.	Improve service management skills	3.00	3.54	2.69	2.92	2.85	2.69
71.	Specialty/tactical training is available equally in all locations	3.31	4.00	2.46	2.85	2.92	2.85
72.	Employees look forward to training to learn something new and "neat" (rather than just fulfilling training days)	2.69	3.38	3.08	3.69	3.38	3.31
73.	Makes new people productive as quickly as possible (faster ramp-up, e.g., to 90% efficiency)	4.38	4.31	4.31	4.08	4.00	4.15
74.	People appreciate product application, not just their technical specialty	3.31	3.23	4.08	3.92	3.31	3.00
75.	Reduced time to market (speed deliverables, reduce design cycle times)	4.62	3.31	4.77	3.69	4.23	4.08
76.	Highlights and contributes to specialized IC (individual contributor) path training requirements	2.85	2.77	3.08	3.23	3.23	3.00
77.	Encourages reuse of information (helps eliminate reinventing the wheel)	3.92	3.46	4.08	3.62	4.08	3.62
78.	Attain training program compliance with standards (e.g., as set by quality councils, ISO, Baldrige, etc.)	2.38	3.23	2.31	2.77	2.54	2.69
79.	Develop high level of general (employee) satisfaction with training program	3.69	4.00	3.38	3.85	3.54	3.00
80.	Improve project planning/management (clarifies project goals)	3.31	3.62	4.23	3.85	3.92	3.38
81.	Attracts/retains key employees (best & brightest); develops employee loyalty	4.08	3.31	4.46	3.54	4.46	2.77
82.	Relationship between (1) advanced technology (2) platform and (3) product development is made seamless	3.31	2.92	3.15	2.23	3.38	2.69
83.	Employees are well-prepared for new (and future) technology	3.92	4.15	4.00	4.00	3.77	3.54
84.	Develop tool knowledge and proficiency required to perform the job	4.00	4.77	3.54	3.77	3.54	3.69
85.	Builds employee capacity and skill set to execute broader and more complex tasks	4.00	4.46	3.85	3.85	3.46	3.31
86.	Swift/immediate knowledge use (less than 3 months after course)	3.46	4.54	3.77	3.69	3.31	2.92
87.	Reduced burden on mentors for training new arrivals	2.38	3.15	2.92	3.15	3.23	3.15
88.	Foster shared, collaborative, team learning in the organization	3.85	3.62	3.46	3.31	3.54	3.31
89.	More productive customer interactions (better value of time spent by both designers & customer)	3.62	2.54	4.23	3.15	3.69	2.77
90.	Develop critical/core expertise (essential scarce skills & depth of knowledge	4.15	4.54	4.23	4.00	3.62	3.69
91.	Improve general product knowledge among employees	3.00	3.77	3.46	3.54	3.38	3.62
92.	More effective management	3.46	3.08	3.85	3.38	3.92	3.31
93.	All designers know & follow [organizational] development processes	3.00	3.62	3.08	2.77	3.54	3.23
94.	Employees get credit/acknowledgment/rewards for their learning achievements	3.31	3.15	2.62	3.23	2.69	1.85
95.	Training as self development (fun) as well as directly making us more productive (useful)	3.08	3.54	3.08	3.23	2.92	2.77
96.	Improve ability to turn product capabilities into value for the customer	3.85	3.38	4.46	3.38	3.62	2.85
97.	Promote usage of in-house expertise (enables less dependency on external expertise)	3.38	3.23	3.00	2.54	2.85	3.08
98.	Provide an increased understanding of roles within a project	2.69	2.38	3.00	2.62	3.31	2.38
99.	Employees perceive that they are receiving the best training available	3.62	4.00	3.54	3.85	3.08	3.23
100.	Integrate training with on-the-job learning (both recognize & support each other)	4.08	4.31	3.38	3.92	3.62	3.77

[a] Notes. Each statement was rated for Training Result Importance (TRI) and Training Evaluation Importance (TEI). The same scale was used for both ratings: 1 = Relatively Unimportant; 2 = Somewhat Important; 3 = Moderately Important; 4 = Very Important; 5 = Extremely Important.

[b] $n = 13$ for each group.

processes, and consequences of training evaluation. The application of various qualitative and quantitative methods, such as interview data and analysis (Maxwell, 1996; Miles and Huberman, 1994) and survey techniques (Hinkin, 1998; Miller, 1994), would serve to complement the results described here.

Acknowledgements

The results reported here are derived from the senior author's doctoral (PhD) dissertation: *Stakeholder Variation in Perceptions About Training Program Evaluation* at the Faculty of Education, University of Ottawa, Canada (completed May, 1999). The research has benefited from input by committee members Swee Goh, Colla MacDonald, Maurice Taylor, and to external member William Trochim. Special thanks are also extended to Stephanie Macaulay for her useful editorial suggestions.

Appendix A

See Table A1.

Appendix B

B.1. Stress values

The stress values obtained for the three concept maps used in the study are as follows: sponsors (0.2474, 11 iterations); participants (0.2882, 12 iterations); providers (0.3045, 12 iterations). According to Concept Systems Inc. (1996b) the idea of the stress value in concept mapping is similar to the idea of the reliability of a measurement. Stress measures the degree to which the distances on the map are discrepant from the values in the input similarity matrix. High stress values imply that there is a greater discrepancy and that the map does not represent the input data as well — low stress values imply a better fit. Some (mainly those who work with extremely well-behaved data like the perception of the similarities of colors or sounds) argue that it is desirable to have a stress value of 0.10 or lower, but this will seldom be attained in concept mapping. However, it should be recognized that their low stress-value expectations are based on experience with much better controlled psychometric testing environments — not usually the case in concept mapping.

In concept mapping, the facilitator should use the stress indicator as a rough guideline of the degree to which the map represents the grouping data. High stress values may imply that there is more complexity in the similarity matrix than can be represented well in two dimensions, that there was considerable variability or noise in the way people grouped the statements, or both. In general, stress values will be lower (i.e., the map will be a better fit) when there are more statements and more people rating the statements than otherwise. A high stress value (i.e., greater than 0.35) may warn the facilitator that there may be some difficulty in interpreting the map sensibly.

B.2. Bridging

A bridging value is computed for each statement and cluster as part of the concept mapping analysis after the concept map is computed. As an index a bridging value always ranges from 0 to 1.

The usefulness of the bridging value is that it indicates whether a statement was sorted with others that are close to it on the map or whether it was sorted with items that are farther away on the map. This index helps in the interpretation of what content is associated with specific areas of the map. For example, statements with lower bridging values are better indicators of the meaning of the part of the map they are located in than statements with higher bridging values.

Bridging can also be computed at the cluster level by taking the average of statement bridging indices in the cluster. Clusters with higher bridging values are more likely to "bridge" between other clusters on the map. Clusters with low bridging values are usually more cohesive, easier to interpret, and reflect the content well in that part of the map. The bridging average (i.e., the arithmetic average of all statement bridging values) for each of the respective concept maps was 0.38 (training providers), 0.50 (training sponsors), and 0.51 (trainees).

References

Abernathy, D. J. (1999). Thinking outside the evaluation box. *Training & Development*, *53*(2), 19–23.

Alkin, M. C. (1991). Evaluation theory development: II. In: M. W. McLaughlin, & D. C. Phillips, *Evaluation and education: At quarter century* (pp. 91–112). Chicago: University of Chicago Press Nineteenth Yearbook of the National Society for the Study of Education, Part II.

Alkin, M. C., Hofstetter, C. H., & Ai, X. (1998). Stakeholder concepts in program evaluation. In: A. Reynolds, & H. Walberg, *Advances in educational productivity*, vol. 7 (pp. 87–113). Greenwich, CT: JAI Press.

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, *42,* 331–342.

Altschuld, J. W., & Zheng, H. Y. (1995). Assessing the effectiveness of research organizations: An examination of multiple approaches. *Evaluation Review*, *19*(2), 197–216.

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

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bassi, L., Benson, G., & Cheney, S. (1996). The top ten trends. *Training & Development*, *50*(11), 29–33.

Brinkerhoff, R. O. (1989). Using evaluation to transform training. In: R. O. Brinkerhoff, *Evaluating training programs in business and industry* (pp. 5–20). San Francisco: Jossey-Bass.

Broad, M. L., & Newstrom, J. W. (1992). *Transfer of training: Action-packed strategies to ensure high payoff from training investments*. Reading, MA: Addison-Wesley.

Brown D.C. (1994). *How managers and training professionals attribute causality for results: Implications for training evaluation*. Unpublished Doctoral Dissertation, College of Education, University of Illinois, Urbana-Champaign.

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science*, *2*(1), 40–57.

Campbell, J. P. (1976). Contributions research can make in understanding organizational effectiveness. In: S. L. Spray, *Organizational effectiveness: Theory, research, and utilization* (pp. 29–45). Kent, OH: Kent State University Press.

Chelimsky, E. (1987). The politics of program evaluation. In: D. S. Cordray, H. S. Bloom, & R. J. Light, *Evaluation practice in Review. New directions for program evaluation*, vol. 34.

Concept Systems Inc. (1996a). *Concept system: Facilitator training manual*. Ithaca: Author.

Concept Systems Inc. [Computer software]. (1996b). Ithaca, NY: Author.

Conference Board Inc. (1997). The value of training in the era of intellectual capital: A research report (Report No. 1199-97-RR). New York: Author.

Connolly, T., Conlon, E. J., & Deutsch, S. J. (1980). Organizational effectiveness: A multiple-constituency approach. *Academy of Management Review*, *5*(2), 211–217.

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, *14*(4), 397–418.

Cousins, J. B., & Earl, L. M. (Eds.) (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer.

Cousins, J. B., & MacDonald, C. J. (1998). Conceptualizing outcomes for the evaluation of management training in technology-based companies: A participatory concept mapping application. *Evaluation and Program Planning*, *21*(3), 333–344.

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

Darrah, C. N. (1995). Workplace training, workplace learning: A case study. *Human Organization*, *54*(1), 31–41.

Drucker, P. F. (1993a). *Post-capitalist society*. New York: Harper Business.

Drucker, P. F. (1993b). Professionals' productivity. *Across the Board*, *30*(9), 50.

Drucker, P. F. (1995). *Managing in a time of great change*. New York: Truman Talley.

Everitt, B. (1980). *Cluster Analysis (2nd)*. New York: Halsted Press/ Wiley.

Fitz-enz, J. (1988). Proving the value of training. *Personnel, March,* 17–23.

Geber, B. (1995). Does your training make a difference? Prove it!. *Training*, *March,* 27–34.

Gill, S. J. (1989). Using evaluation to build commitment to training. In: R. O. Brinkerhoff, *Evaluating training programs in business and industry* (pp. 35–43). San Francisco: Jossey-Bass.

Gold, N. (1983). Stakeholders and program evaluations: Characterizations and reflections. In: A. S. Bryk, *Stakeholder-based evaluation. New directions for program evaluation*, vol. 17 (pp. 63–72). San Francisco: Jossey-Bass.

Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park: Sage.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121.

Kearsley, G. (1982). *Costs, benefits, and productivity in training systems*. Reading, MA: Addison.

Kirkpatrick, D. L. (Ed.) (1975). *Evaluating training programs*. Madison, WI: ASTD.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco: Berrett-Koehler.

Kirkpatrick, D. L. (1996). Great ideas revisited. *Training & Development*, *50*(1), 54–59.

Kolb, D. G., & Shepherd, D. M. (1997). Concept mapping organizational cultures. *Journal of Management Inquiry*, *6*(4), 282–295.

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

Lewis, T. (1996). A model for thinking about the evaluation of training. *Performance Improvement Quarterly*, *9*(1), 3–22.

Likert, R. (1977). Measuring organizational performance. In: S. J. Carroll Jr, F. T. Paine, & J. B. Miner, *The management process: Cases and readings* (pp. 316–332). New York: Macmillan (Reprinted from Harvard Business Review, March–April, 1958).

Love, A. J. (1991). *Internal evaluation: Building organizations from within*. Newbury PK: Sage.

Mark, M. M., & Shotland, R. L. (1985). Stakeholder-based evaluation and value judgments. *Evaluation Review*, *9*(5), 605–626.

Maxwell, J. A. (1996). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.

McLinden, D. J. (1995). Proof, evidence, and complexity: Understanding the impact of training and development in business. *Performance Improvement Quarterly*, *8*(3), 3–18.

McLinden, D. J., & Trochim, W. (1998). Getting to parallel: Assessing the return on expectations of training. *Performance Improvement*, *37*(8), 21–26.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*, (2nd ed.). Thousand Oaks, CA: Sage.

Miller, T. I. (1994). Designing and conducting surveys. In: J. S. Wholey, H. P. Hatry, & K. E. Newcomer, *Handbook of practical program evaluation* (pp. 271–292). San Francisco: Jossey-Bass.

Mintzberg, H. (1983). *Power in and around organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Moad, J. (1995). Calculating the real benefit of training. *Datamation*, *April,* 45–47.

Nadler, L. (1984). Human resource development. In: L. Nadler, *Handbook of human resource development* (pp. 1.1–1.47). New York: Wiley.

Palumbo, D. J. (Ed.) (1987). *The politics of program evaluation*. Newbury Park: Sage.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text*, (3rd ed.). Thousand Oaks, CA: Sage.

Pepitone, J. S. (1995). *Future training*. Dallas: AddVantage Learning Press.

Pfeffer, J. (1981). *Power in organizations*. Marshfield, MA: Pitman.

Pfeffer, J. (1994). *Competitive advantage through people: Unleashing the power of the work force*. Boston: Harvard Business School Press.

Phillips, J. J. (1996). ROI: The search for best practices. *Training & Development*, *50*(2), 42–47.

Phillips, J. J. (1991). *Handbook of training evaluation and measurement methods*, (2nd ed.). Houston: Gulf.

Rizzo-Michelin L.L. (1998). *Concept Mapping in Evaluation Practice and Theory: A Synthesis of Current Empirical Research*. Unpublished master's thesis, University of Ottawa, Canada.

Setze, R. (1997). Performance needs assessment in a corporate setting. In: *Paper presented at the meeting of the American Evaluation Association, November, San Diego, CA*.

Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory research and practice since 1986. *Evaluation Practice*, *18*(3), 195–208.

Spray, S. L. (1976). Organizational effectiveness: Theory, research, and utilization introduction. In: S. L. Spray, *Organizational effectiveness: Theory, research, and utilization* (pp. 1–4). Kent, OH: Kent State University Press.

Stake, R. E. (1983). Stakeholder influence in the evaluation of cities-in schools. In: A. S. Bryk, *Stakeholder-based evaluation. New directions for program evaluation*, vol. 17 (pp. 15–30). San Francisco: Jossey-Bass.

Stewart, T. A. (1999). *Intellectual capital: The new wealth of organizations*. New York: Currency Doubleday.

Sumanth, D. J., Omachonu, V. K., & Beruvides, M. G. (1990). A review of the state-of-the-art research on white-collar/knowledge worker productivity. *International Journal of Technology Management*, *5*, 337–355.

Tesoro, F. (1998). Implementing an ROI measurement process at Dell computer. *Performance Improvement Quarterly*, *11*(4), 103–114.

Thompson, J. D. (1967). *Organizations in action*. New York: McGraw-Hill.

Toffler, A. (1990). *Powershift: Knowledge, wealth and violence at the edge of the 21st century*. New York: Bantam.

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

Trochim, W. (Ed.) (1989a). Concept mapping for evaluation and planning (special issue). *Evaluation and Program Planning*, *12*(1) Special Issue.

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

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

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

Trochim, W. (1990). Pattern matching and program theory. In: H. C. Chen, *Theory-driven evaluation. New directions for program evaluation*. San Francisco, CA: Jossey-Bass.

Trochim, W. (1993). The reliability of concept mapping. In: *Paper presented at the Annual Meeting of the American Evaluation Association, Dallas, TX., November*.

Weiss, C. H. (1983). The stakeholder approach to evaluation: origins and promise. In: A. S. Bryk, *Stakeholder-based evaluation. New directions for program evaluation*, vol. 17 (pp. 3–14). San Francisco: Jossey-Bass.

Weiss, C. H. (1987). Where politics and evaluation research meet. In: D. J. Palumbo, *The politics of program evaluation* (pp. 47–70). Newbury Park: Sage.