From Puzzles to Problems: Assessing the Value of Education in a Business Context With Concept Mapping and Pattern Matching

Andersen Consulting

Daniel J. McLinden and William M.K. Trochim

Determining the value of an investment in education is a critical issue in business today. Perhaps equally important is the need to apply sound methodology to assess the value of programs designed to effect a change in people and their organizations. Although a variety of social science research methods are currently in use in efforts to determine the value of education, there is a need to expand the evaluation tool kit. In addition to assessing program results, this case illustrates and recommends a method of rigorously conceptualizing program impact. We have used this methodology—concept mapping and pattern matching—to conceptualize program impact and to then provide a framework for determining the impact of a program on expected results. Results from this case showed that the program studied successfully achieved the performance expected from training, although the outcome diminished slightly over time. In addition to this overall assessment of effectiveness, the method provided detailed diagnostic information about which topics were being conveyed well (linkage) and which were failing to meet the expected outcome (disconnect). This information was especially valuable for development teams that needed to know how best to redesign training programs as part of a continual improvement effort.

Background

What value does business derive from its investment in education? Business executives base decisions about the operation of a business on the answer to that question, and at the same, training executives

This case was prepared to serve as a basis for discussion rather than to illustrate either effective or ineffective administrative and management practices.

attempt to provide the answer. Although the need to assess and communicate the value of education seems clear, a comprehensive response to this need has not been forthcoming. The level of evaluation in the field of training seems small compared with the need. The scope, at least for the investment in training in the United States, is estimated in excess of $50 billion. Obviously the figure is larger if we take an international view. Clearly, the value of this level of investment needs to be substantiated to be maintained, and quite possibly this level even needs to be increased. However, neither action will occur or endure for long without compelling evidence of value, and efforts to demonstrate the value of training need to grow.

When compelling evidence of training value is required, often the quick reaction is to calculate the costs of training and the financial benefits that result from the investment in training. This approach to evaluation applies in educational programs the analytical techniques that are similar to those used to evaluate stocks, bonds, and other financial variables. For example, placing money in various financial instruments clearly leads to some measurable outcome. Through judicious insight, an investor may have a return that is both positive and in excess of other competing alternatives for the same investment dollars. By applying various formulas, the investor can clearly see the outcome of an investment strategy. This evaluation approach is compelling because it holds promise for determining the monetary value of an investment in education. To be sure, when feasible, the capability to apply the same techniques to training provides strong estimates of the relationship between training costs and training value. However, the difficulty that arises is in the level of proof necessary to claim the benefit (or loss) that can be attributed to training. The variables involved in a financial transaction are straightforward, and the action of investing in specific instruments is linked unequivocally to the outcome. However, education involves a series of human transactions that introduce ambiguity. Disentangling the variables that introduce ambiguity to achieve a level of scientific proof is difficult in most cases and arguably not cost-effective in others. As a result, some programs may be able to demonstrate economic gains. Other programs may be unable to demonstrate gains, not because the program is at fault but because the unique features that make the proof of a financial result possible are not present. Such features include control over the implementation of the program, access to financial measures of outcomes, and a clear relationship between the training program and any changes in business activity. Therefore, what of situations where the

cost of experimental methods is too high relative to the benefit or where it is simply not possible to implement a financial analysis or where other evidence would suffice? In part, the difficulty of addressing the question may be that the definition of educational value has become synonymous with the analysis of work hours, profits, and so forth. To be sure, these variables are important, but this view may be detrimental if it stymies evaluation based on other models of how the educational process unfolds in and adds value to an organization. Equally compelling methods to demonstrate value are necessary.

What are the alternatives? Adding to the evaluation tool set may first involve a shift in assumptions about the nature of training evaluation.[1] Specifically, evaluating training should be viewed not as a puzzle but as a complex problem. Consider the difference between solving a puzzle and solving a problem. Completing a puzzle requires one to fit all the pieces to complete the picture or in terms of a financial model, to calculate all the financial formulas to determine value. Although solving a puzzle can indeed be complex, the process assumes that all the pieces are given and ultimately a single clear and unambiguous answer exists. In other words, costs can be specified, revenues to training determined, and a clear formulaic relationship exists between the two variables. Contrast puzzle solving with the problem solving typical of work environments. Multiple "right" answers generally exist for business problems, establishing "truth" may involve making trade-offs between competing alternatives, and less-than-ideal "solutions" are sometimes selected due to a lack of resources or a host of other constraints. The domain of education is no less complex.

Rather than completing a puzzle, assessing educational value is an effort in solving problems in which the training professional must, in effect, create some of the pieces of the puzzle, act without knowing how the final picture will look, and be satisfied when the final picture contains some ambiguity. This view of educational value is in contrast to a view of training evaluation as solving a puzzle or, in other words, as applying measurement tools. Viewing the determination of training value as a measurement exercise is problematic because training value is not an objective entity existing somewhere out there and waiting to be measured. Rather, value represents a constructed reality that depends on the views of individuals or stakeholder groups associated with the process. Consequently, the construction of the meaning of value cannot come solely from an analysis of business activities but must include the constructed interpretations of value from multiple stakeholders. Over the course of several projects

designed to answer the question, "What is the value of this program?" it became apparent that educational evaluation must encompass methods that can deal with the psychological and sociological complexity of the educational enterprise. Such methods need to be able both to measure value in the development and delivery of a complex service and communicate clearly the results to business people not necessarily interested in the fine points of complex statistics and research design.

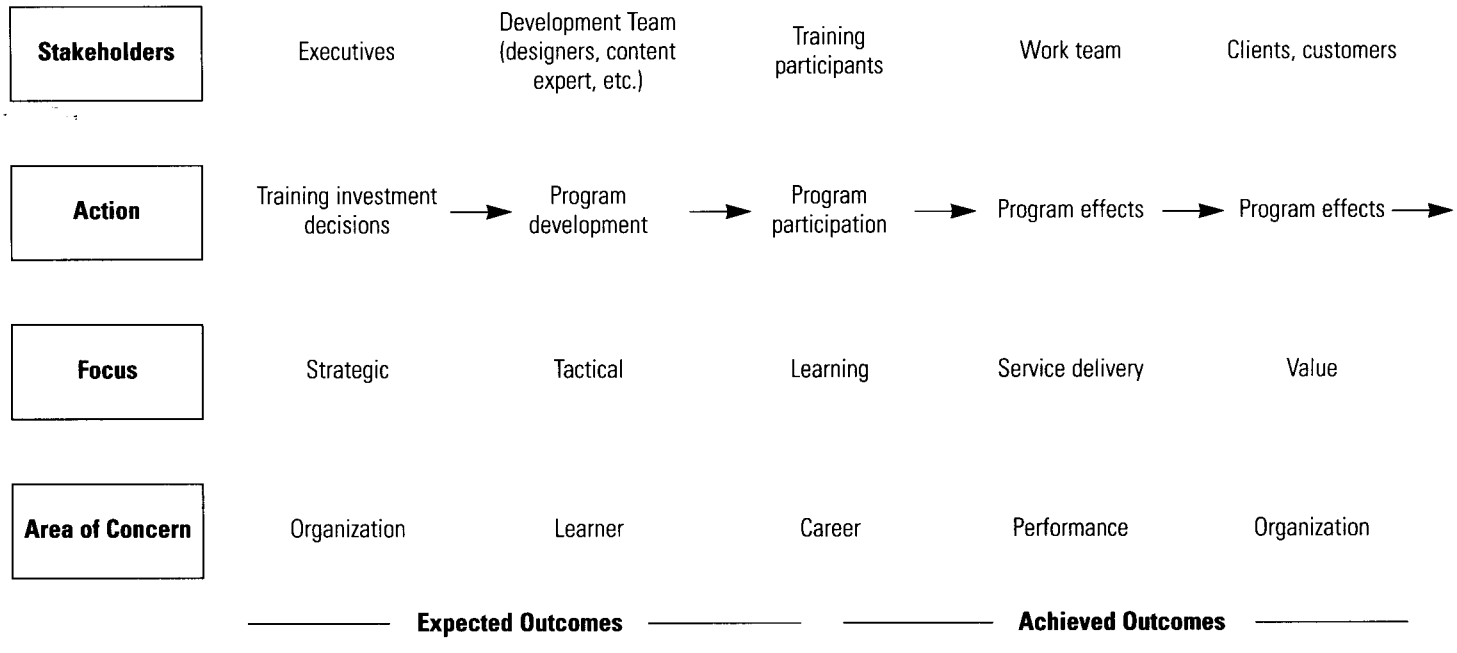
Stakeholder Transactions in Educational Evaluation

Much of the effort in determining value seems devoted to addressing the question, "To what extent did the N days spent at the XYZ program lead to increased sales?" This is a focus on the educational event as a cause of a specific economic result. A different premise and one that forms the basis for an alternative approach to evaluation is to focus not on the educational event per se, but on the human transactions involved in the educational cycle. Figure 1 provides a generic view of those transactions, some of which correspond to the educational event whereas others precede and follow. There are two key aspects portrayed in this figure.

First, assessing program effects (see the far right side of the second row in figure 1) represents one link in a chain of events. The chain begins with the individuals who have decided to fund an educational response to some business need. The distance between the strategic business decision and the ultimate effects for the client or customer is populated with other stakeholder groups. It would be overly idealistic to assume that developers develop what they are directed to build, learners learn what they are told are the objectives, and supervisors in the field support the use of skills that executives thought were necessary. Most organizations are loosely coupled, and each stakeholder group will bring a different set of concerns to the educational process and will, in turn, influence that process in different ways.

For example, business executives are typically concerned with developing and implementing strategies to be responsive to the marketplace; they may conclude that training is one way to develop or support a response to market forces. The instructional development team, although not unconcerned with the marketplace, focuses its decision making on techniques for delivering instructional content. It is not that an instructional development team will work at cross-purposes to the strategic intents of an executive team, but its focus will be different. The developers' mental model of the program will

Figure 1. A transactional model of educational value.



drive day-to-day decisions about building a product. That focus may not be an exact replica of the executives' strategically focused mental model of the final product. Similarly, although the executive team's primary concern is strategy, its members have at least implicit tactical beliefs about the program, such as how it will be delivered and the types of activities. The result is that the arrows in figure 1 can portray either a potential linkage or a disconnect between stakeholders' groups. That is, the implicit mental models concerning strategy and tactics that any stakeholder group may hold vary between being quite similar (a linkage) to quite dissimilar (a disconnect) when compared to those of another stakeholder group. If stakeholders do not agree on a rigorous and explicit definition of the basis for evaluating the program, it is likely that the program cannot be evaluated, regardless of the methods used. Consequently, a basic premise of determining value should be that, the *evaluation effort must not only encompass post-training evaluation but also begin prior to the development of training. Further, the evaluation must be concerned with making explicit and then integrating the expectations of stakeholders across the lifecycle.*

Second, there is a dichotomy in figure 1 between those who play a dominant role in setting expectations and those who are the recipients of programs designed on the basis of those expectations. Mentally divide figure 1 in half vertically so that "executives" and "development team" appear on the left and the remaining stakeholder groups are on the right. On the left side, the strategic and tactical stakeholders will make decisions on the basis of what they want to accomplish, in other words, they will set the expectations for the program. On the right side are the stakeholders who will experience the effects of the program. Using an algebraic equation as a metaphor, the outcomes from a program ought to equal the intents of the program. In cases in which intents and outcomes do not match, an inequality exists. This inequality can be remedied by working to change one side of the equation or the other. In other words, modify the intents of the program or modify the program to achieve the intended outcomes. Therefore, a second premise is that *the value of a program ought to be judged on the basis of the degree of linkage between expected and achieved outcomes.*

In light of this view of the educational process, the methodological process becomes one of evaluating the linkages between the human transactions. Although the training event is certainly part of the transactions, the focus has to be on the human transactions involved in program development and delivery. The question then becomes, "How can these transactions be evaluated to determine the level of train-

ing value?" The methodology of concept mapping and pattern matching holds promise for addressing this need to value the return on the investment in education on the basis of this transactional model of the educational process.

An Example of a Concept Mapping and Pattern Matching Application

Procedures

The authors began implementation of this evaluation methodology with a school designed to develop new programming skills for experienced programmers. The school was one week long and relied extensively on a team-based approach to learning. Concept mapping was accomplished in four steps: (1) generate expectations, (2) sort the expectations, (3) rate the expectations, and (4) analyze the results and facilitate dialogue. Readers seeking a more detailed explanation of the methodology of concept mapping and pattern matching should read the articles by William Trochim listed at the conclusion of this chapter.

The training development team generated 66 statements that described the outcomes of the training program. These were detailed and specific statements of what a training participant will know or be able to do as a result of successful completion of the program. Next, developers, managers, and content experts were asked to complete a process that made explicit their mental model of the program and intended effects. Individuals worked alone and completed two tasks.

First, they sorted or organized each of the 66 outcomes into concepts. For the sorting task, statements were printed on individual cards[2] along with an identifying number. Nine participants were instructed to work individually and group the 66 statement cards into groups "in a way that makes sense to you." The only restrictions in this sorting task were that there could not be (a) 66 groups of one item each, (b) one group consisting of all 66 items, or (c) a "miscellaneous" group with multiple items (any item thought to be unique was to be put in its own separate group). After sorting the statements, each participant recorded the contents of each group by listing the statement identifying numbers along with a short label describing the contents of the group. In effect each stakeholder was describing his or her unique view of the structure of these training outcomes.

Second, individuals rated the degree of impact each outcome should have in the program. In the rating task, the statements were listed in questionnaire form, and each stakeholder was asked to rate each

statement on a 5-point Likert-type response scale in terms of how important the statement is to the training program (1=relatively unimportant, and 5=extremely important). All of the statements had some degree of importance with respect to the program. Therefore, it was stressed that the rating should be considered a relative judgment of the importance of each item to all the other items brainstormed. As a consequence, a rating of one or relatively unimportant did not mean that an item was not important but was simply relatively less important than other items that represented other aspects of the training program. In this step each stakeholder was able to provide his or her unique perspective on the value associated with various aspects of the training program.

Analysis

The analytical task combined the unique and individually generated data from sorting and rating to create a group perspective of the training program. Aggregating the data generated in this process requires use of sophisticated statistical algorithms, multidimensional scaling, and cluster analysis as well as unique graphic portrayals of the results. Our use of an integrated software program[3] kept the statistical aspects of this process transparent to the participants in this process. However, we include some of the details here to provide the reader with a sense of the underlying rigor in this process. The concept mapping analysis begins with construction from the sort information of an NxN binary, symmetric matrix of similarities, Xij. For any two items i and j, a 1 was placed in Xij if the participant placed the two items in the same group; otherwise a 0 was entered. The total NxN similarity matrix, Tij, was obtained by summing across the individual Xij matrices. Thus, any cell in this matrix could take integer values between 0 and 9 (that is, the nine people who sorted the statements); the value indicates the number of people who placed the i,j pair in the same group. The total similarity matrix Tij was analyzed using nonmetric multidimensional scaling (MDS) analysis with a two-dimensional solution. The analysis yielded a two-dimensional (x,y) configuration of the set of statements based on the criterion that statements grouped together most often are located more proximately in two-dimensional space, whereas those grouped together less frequently are further apart.

The x and y coordinates for each item were the input for hierarchical cluster analysis.[4] Using the x and y coordinates as input to the cluster analysis in effect forces the cluster analysis to partition

the items into nonoverlapping clusters. There is no simple mathematical criterion by which a final number of clusters can be selected. The procedure followed here was to examine an initial cluster solution that, on average, placed five statements in each cluster. Then, successively lower and higher cluster solutions were examined, with a judgment made at each level about whether the merger or split of clusters seemed substantively reasonable. In this case, the seven-cluster solution preserved the most detail and yielded substantively interpretable clusters of statements.

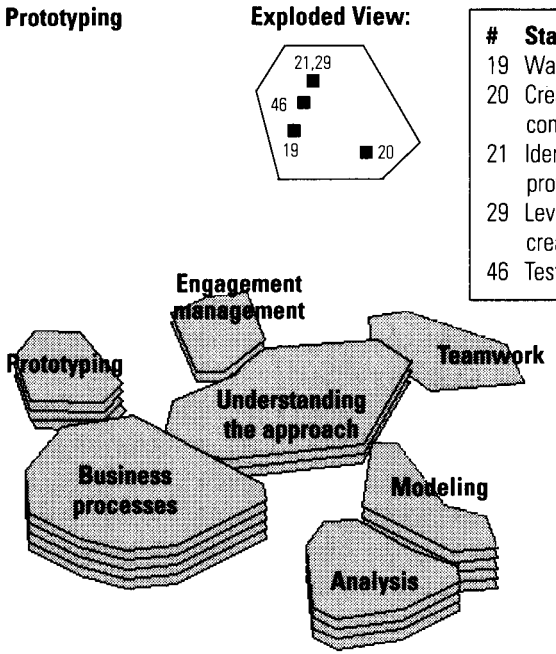
Concept Mapping Results

The x and y coordinates of the 66 points were graphed in two dimensions. This point map displayed the location of all the training outcomes on the basis of the stakeholders' perspectives. Outcomes closer to each other indicated that stakeholders' meanings were similar, whereas those that were far apart were quite different.

A "cluster map" was also generated, which partitioned the original 66 points enclosed by boundaries of seven clusters, or concepts. Finally, the rating data were averaged across persons for each item and each cluster. This rating information was depicted graphically in a cluster rating map, which showed the cluster average rating using the third dimension (higher = more important). The evaluators then presented these results to the participant group and worked with them to identify a short text label indicative of the content for each of the seven clusters.

The result of the effort to set expectations was a map of the seven clusters of outcomes, as shown in figure 2. The concepts that emerged (for example, teamwork, prototype development, analysis[5]) represented the interpretation of the specific ideas in each of these clusters. The location of each of the 66 separate outcomes (that is, the point map) is not shown because that level of detail is beyond the scope of this paper. However, as an example, the specific outcomes that clustered into a concept termed prototyping are shown. A similar level of detail exists within each of the clusters. Finally, the relative importance of each of the concepts is indicated by the height (higher = more important). The height of each cluster results from the average importance value of each of the items contained in a cluster. These values resulted from the relative ratings of each stakeholder involved in this task. In effect, this concept map portrays the intent of this school as expressed by the combined expectations from multiple stakeholders.

Figure 2. Concept map of expected outcomes.



Pattern Matching Results

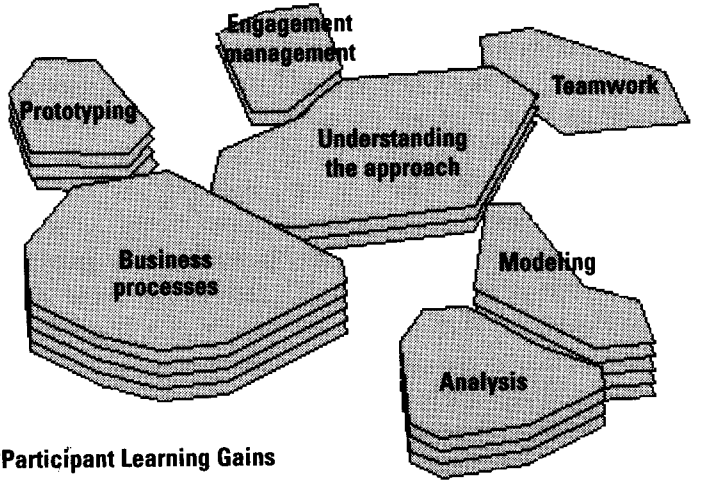
Assuming that the program's participants ought to be affected in a way that is consistent with the intent of the school (that is, the importance of each component), we obtained participant data on each of the 66 outcomes at the educational events. Although data could have been obtained in a variety of ways, we decided to directly ask each of the participants to indicate their level of learning on each of the 66 outcomes. At the initial test of the program (alpha test), five participants assessed their competence at the beginning and at the conclusion of the school.

Pattern matching assumes that a correspondence ought to exist between the intended emphases of the program and the observed gains. If the design of the program had the intended effects, the greatest gains in learning ought to be in those areas that were most emphasized (that is, most important) in the program (see figure 3). A visual comparison of the map in figure 2 (expectations based on program design) with the map in figure 3 (outcomes based on learning) shows that for the most part the patterns do match. Concepts that

were more important (that is, had greater height) in the educational design (expectations) showed greater gains (that is, had greater height) in learning on the outcome map. The match suggests that the program was effective in achieving its aims. However, the match also showed several areas were not as closely aligned as desired. Analysis showed a decline and understanding the approach showed a gain. This information provided the development team with direction for work prior to additional testing.

In addition to the quick visual comparison of figures 2 and 3, a more exacting comparison was also used. In figure 4 the average value of each concept was plotted along an axis for expectations and along an axis for outcomes. These two axes were then placed in parallel and lines were drawn to connect corresponding clusters. In cases where relative outcomes matched expectations (linkage), the connecting lines were horizontal or nearly so. If all concepts were in alignment, the graph would resemble a ladder; we have coined the term *ladder graph* to describe these graphs. Likewise, lines crossing at steep angles portray a lack of linkage or relative disconnect between concepts. For example, two characteristics of the pattern in figure 4 are apparent. The outcome area understanding the approach had a stronger gain than was expected, and although analysis was an important area in

Figure 3. Impact of the program on learning outcomes.



the design, this concept did not show a commensurate gain on the outcome side. Rather than a summary grade on performance, this type of analysis is more diagnostic and offers the development team a focus for dialogue about needed continuous improvements.

Subsequent evaluation efforts were aimed at obtaining feedback from participants three to six months after training, as shown in figure 5. Again we elected to ask past program participants (n=23) about their experience. These results showed that although program was generally effective, the match declined in several areas. Former participants did not feel as competent in the key technical skills that were desired from the course. Interviews suggested that the increased disconnects resulted from a lack of opportunities to apply newly learned skills. Additionally, the most notable change was the teamwork cluster, which suggests that this concept or skill area may be more important to job performance than expected at the outset of the program.

Implications of This Example for Program Development and Evaluation

There are several immediately apparent advantages to using the concept mapping and pattern matching approach to assessing the effects of a training program. First, the process involves all of the relevant stakeholder groups (for example, initiating executives, program developers, managers) in contributing input relevant to their perspectives. Second, because concept mapping is done when the training effort is just beginning, and is followed (through pattern matching) all the way through to outcome assessment, the process assures that there will be continuity from conception to outcome and consensus over time across the various stakeholder groups. Third, because the major products of the process are visual, they are readily understandable by the various stakeholder groups. Developers and initiators are more easily able to recall the important emphases in the training and how these are interrelated. They can easily summarize how the training is affecting outcomes, in specific topical areas as well as for the training program as a whole. Fourth, because the process is based on state-of-the-art statistical methods (although these are transparent to the participating stakeholders), there is a high degree of rigor and credibility in the results. Fifth, the results are useful both for providing detailed diagnostic feedback that suggests to development teams where improvements might be made and providing an assessment of the overall strength of the program effect. For example, it was clear from the initial concept mapping that although analysis was

Figure 4. Pattern match of expectations to learning outcomes.

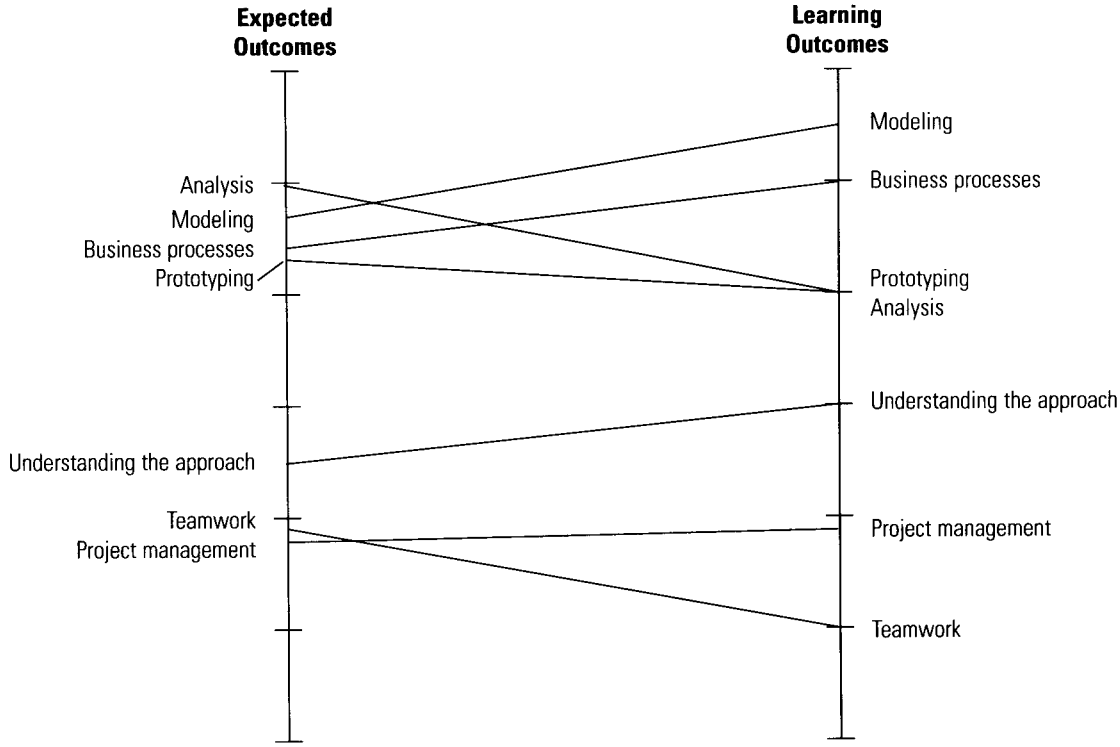
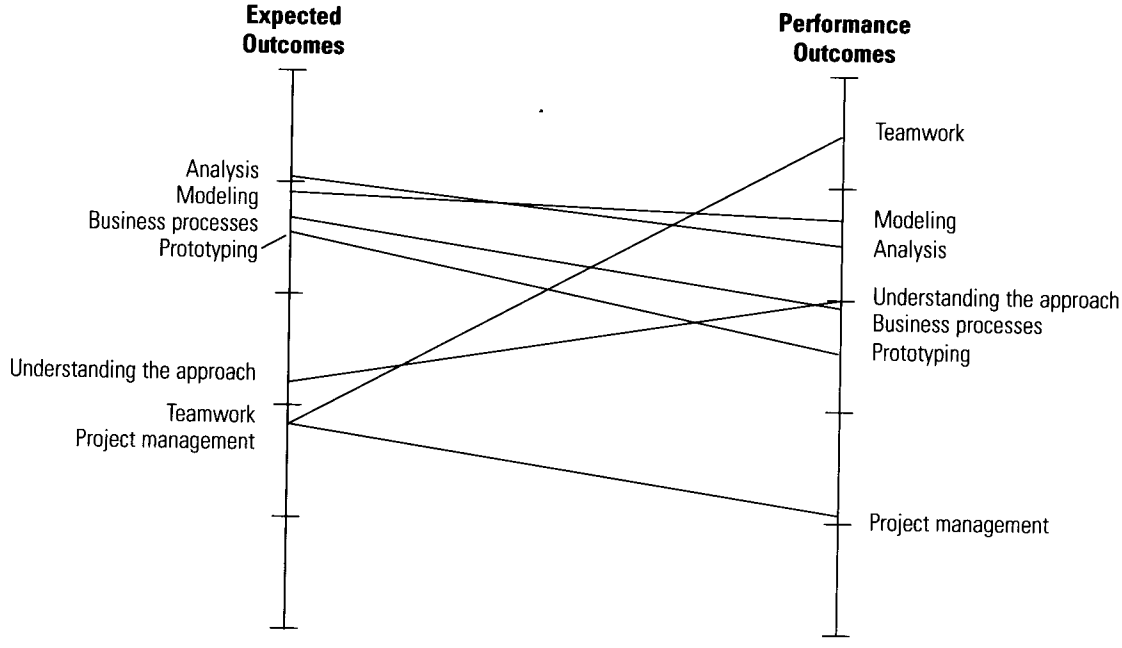


Figure 5. Pattern match of program design to job-related outcomes.



deemed the most important expectation for training, it ranked only fourth in importance in terms of observed gains. This suggested to the development team that they revisit how they were presenting material related to that topic, with the possibility that they might make it more salient in the next session of the training. Conversely, the evidence in the follow-up data indicated that the greatest long-term gains were in the area of teamwork. Although teamwork was rated next to last in importance, developers suggested that the current emphasis on teamwork in the course could be softened (while simultaneously adding emphasis in other more important areas) without any significant loss of desired outcome. At a more macro level, the linkages between the expected and observed outcomes suggest that on the whole the training was getting across the desired material.

Implications of This Methodology for Evaluating the Value of Educational Programs

The seminal points on evaluating the effects of training programs were made by Kirkpatrick, who introduced the concept of levels of evaluation. Kirkpatrick defined four levels: reaction, the degree to which a student liked a program; learning, the degree to which a student gained knowledge, skills, or changed attitudes; behavior, the individual's application of skills on the job; and results, improvement or decreasing problems for the organization. To say that Kirkpatrick's concept of levels resonated with the training community would be an understatement. Discussion of evaluative activities in training is often framed in terms of specific levels. In spite of the fact that many authors recommend moderate use of experimental methods, there is a strong interest in the field in demonstrating value by scientifically proving the economic return from training programs. However, not all programs are amenable to achieving the level of proof implied by experimental approaches. Now may be an opportune time to reconsider the definition of training value along with the utility of all levels, including Level 1.

First, the implicit meaning in levels is that higher levels of evaluation produce a more substantive conclusion than do lower levels. This implication has helped move the training community and our stakeholders away from quick reactions by students as a measure of value. However, it may have also moved the community away from gaining the proper perspective of those involved in the educational enterprise. In some cases, gaining the personal perspective of training participants can have more value than other measures. For example,

in the transactional model, our concern is seeking input from training participants on their specific learning experiences. This is unlike asking students to react to features of the training event (for example, overall quality, timing, instructional materials) about which they would certainly have opinions but would lack the appropriate context to make a properly informed judgment. Training participants are not aware of the management decisions made during the development of the program. Some of these decisions may reflect trade-offs between the desire for optimization and the need to develop and deliver a product within a specified time at a specified cost. Further, a rigorous testing program could have been instituted to establish more valid measures of learning than self-report. However, the time and effort required would be high, and the additional value, as compared to self-report, low. As a result, students' self-assessment of their learning at the conclusion of the event was sufficient evidence to credibly demonstrate the value of the program.

Second, the need for Level 4 evaluation may be driving the field of training toward a focus on assessment strategies that assume that value can be proved through a measurement exercise. Evaluation methods must be responsive to a broad range of issues caused by the increasing complexity of the educational process. In business and industry, education is undergoing substantial change. For example, development teams that used to choose between lectures and self-study now must consider interactive multimedia, team-based learning, integrated performance support, and so on as well as the traditional approaches. Likewise, the business environment is undergoing radical change. *Customer-driven, reengineering, downsizing,* and *teaming* are just a few of the terms describing organizational changes businesses are using to respond to the marketplace. Within this complex and turbulent environment, the traditional approaches to evaluating the value of training also need to undergo some transformation. The methodology of concept mapping and pattern matching seems to have addressed the need to be responsive to the complexity of education in a business environment in a way that portrays evidence (or lack thereof) in a compelling manner. Furthermore, our experience is that this process has the potential to include other types of measures and rely on financial or other measures of organizational productivity to assess outcomes. That is, if the outcomes for concepts could be operationalized as objective measures (money, time, and so on), the pattern match between these outcomes and the expected impact of the program could be evaluated. Likewise, the use of the participants'

reactions to a program also ought to be considered a valid assessment of the educational program. However, as we have demonstrated here, their reactions and those of other stakeholders' groups (for example, supervisors) should be given serious consideration if the goal is a compelling portrayal of value.

Conclusion

Although much of the effort in development and delivery of training programs seems justified and effective on an intuitive level, evidence to substantiate this intuition is, at best, sparse. And with the annual cost of training estimated at tens of billions of dollars, it should come as no surprise that the term *return-on-investment* has become the current hot topic in training. More surprising is the fact that the field has come this far without substantial evidence of what occurs as a result of an investment in education. Although this trend seems to be changing, much of the effort seems aimed at expanding the application of traditional research solutions such as experimentation. Although these traditional evaluation approaches will continue to have value for specific reasons on specific programs, addressing the scope of a multibillion-dollar figure or the value of multiple educational programs across a diverse organization will require more expansive thinking about the techniques of evaluation. In our work at Andersen Consulting Education we believe we have begun the process of redefinition and have successfully expanded methods to address training value. Perhaps most important, the concept mapping and pattern matching approach moves us away from the mentality of solving precut puzzles and toward the idea of viewing the training endeavor as a complex problem involving multiple stakeholder groups who bring different perspectives to the educational effort. Although not abandoning the goal of understanding how training affects the bottom line, the concept mapping and pattern matching approach rightly focuses our attention on achieving the return on our expectations that we had in mind when originally undertaking training development.

Questions for Discussion

1. The case suggests that evaluation of training programs requires both measures and the collaboration of stakeholders. In what situations, other than training, does this model seem appropriate?
2. If the evaluation specialists had been required to draw return-on-investment conclusions, how could the methodology used in this case have been utilized to encompass that need?

3. The results in this case are portrayed graphically. Although the graphs are data driven, tables of figures are not presented. How is this similar or different from other results you have seen? What benefits can you see in this approach to reporting?
4. The case emphasizes conceptualizing program impact through concept mapping. How does this differ from other evaluation efforts with which you are familiar?

The Authors

Daniel J. McLinden is director of evaluation and performance measurement for Andersen Consulting at the Center for Professional Education in St. Charles, Illinois. Over the course of his career in human resource development in business, his consulting efforts have included program evaluation, needs assessment, testing, and organizational assessment. His current focus is on conceptualizing program impact and the evaluation of investments in the development of human resources. Much of his current work in writing and presentations has focused on the importance of linking program conceptualization and evaluation. McLinden holds a doctorate in educational psychology from Northern Illinois University. He can be contacted at the following address: Andersen Consulting, Center for Professional Education, 1405 North Fifth Avenue, St. Charles, IL 60174-1264; e-mail: daniel.j.mclinden@ac.com.

William M. K. Trochim is a Professor at Cornell University and the creator of the concept mapping and pattern matching methods, the programmer of The Concept System computer program, and a co-founder of Concept Systems, Inc. He received his Ph.D. in methodology and evaluation research from the Department of Psychology at Northwestern University. He specializes in program evaluation and planning at Cornell, is a faculty member in the Department of Human Service Studies, and is a member of the graduate fields of education, epidemiology, and human development and family studies. Trochim has published widely on evaluation and social research methods, and is an internationally recognized expert in evaluation methods, including conceptualization approaches, causal assessment, experimental design, and the theory of validity in research.

Resources and References

Statistics and Research Methodology

Everitt, B. (1980). *Cluster Analysis.* New York: Halsted Press.

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

Rosenberg, S., and M.P. Kim. (1975). "The Method of Sorting as a Data Gathering Procedure in Multivariate Research." *Multivariate Behavioral Research, 10,* 489–502.

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

Training Evaluation and Impact Assessment

Kirkpatrick, D.L. (1987). "Evaluation." In R.L. Craig and L.R. Bittell (editors), *Training and Development Handbook* (pp. 301–319). New York: McGraw-Hill.

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., M.J. Davis, and D.E. Sheriff. (1993). "Impact on Financial Productivity: A Study of Training Effects on Consulting Services." *Human Resource Development Quarterly. 4*(4), 367–375.

Phillips, J.J. (1994). *Return on Investment in Human Resource Development: Cases on the Economic Benefits of HRD.* Alexandria, VA: American Society for Training & Development.

Phillips, J.J. (1997). *Return on Investment in Training and Performance Improvement Programs.* Houston: Gulf.

Levels of Training Investments

Training Budgets. (1995). "1995 Industry Report." *Training, 32*(10) 41–48.

Training Budgets. (1996). "1996 Industry Report." *Training, 33*(10) 41–49.

Concept Mapping and Pattern Matching

Concept mapping software: Concept Systems, Inc. 118 Prospect Street, Suite 309, Ithaca, NY 14850; 607.272.1206.

Trochim, W.M.K. (1989a). "An Introduction to Concept Mapping for Planning and Evaluation." *Evaluation and Program Planning, 12,* 1–16.

Trochim, W.M.K. (1989b). "Outcome Pattern Matching and Program Theory." *Evaluation and Program Planning, 12,* 355–366.

Notes

[1] Readers interested in additional research methods should consult Phillips (1997) for methods of estimating financial variables.

[2] At the time this project was conducted, we used the paper version of the process described here. Subsequently, we have used the Windows version of Concept Systems© Inc. software for other evaluation projects and have, for the most part, utilized an online version of the sorting and rating tasks contained in the software.

[3]Results reported here used concept mapping software by Concept Systems© Inc. Readers interested in additional information on the software should contact the software company. Readers interested in the statistical processes should consult the suggested readings at the end of this chapter.

[4]Concept Systems© software uses Ward's method as the basis for defining a cluster

[5]Due to the proprietary nature of the content, the specific technical skill areas are not described completely.