Getting to Parallel: Assessing the Return on Expectations of Training

by Daniel McLinden and William M.K. Trochim

Dan McLinden directs the Evaluation & Performance Measurement Team 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. Dr. McLinden's 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. He may be contacted by mail: Andersen Consulting, Center for Professional Education, 1405 North Fifth Avenue, St. Charles, IL 60174-1264; or by email: daniel.j.mclinden@ac.com.

William M.K. Trochim is a professor at Cornell University in the field of Program Evaluation and Planning in the Department of Policy Analysis and Management. Dr. Trochim is also the creator of the concept-mapping and pattern-matching methods, the programmer of The Concept System computer program, and a cofounder of Concept Systems, Inc. He may be contacted by phone: (607) 272-1206; by mail: Room 132 MVR Hall, Cornell University, Ithaca, NY 14853 or Concept Systems Inc., 118 Prospect Street, Suite 309, Ithaca, NY 14850; or by email: wmt1@cornell.edu.

Increasingly, "What value are you adding?" is the type of scrutiny training development managers face. Effectively responding to this increase in scrutiny requires a process with sufficient rigor to provide credible evidence of value. In our work, we also recognized the need for the process to be repeatable as well as rigorous. Rigorous because we wanted to be able to make strong statements about program impact and to withstand challenges to those assertions. Repeatable because we wanted to avoid reinventing the measurement strategy for each program. However, we also recognized the need to balance rigor and repeatability with the uniqueness of each training product.

Our approach to address the demands of rigor, repeatability, and uniqueness consisted of two key features: 1) obtain a preponderance of evidence demonstrating that the expectations for program impact originally leading to the investment were met and 2) collect evidence with sufficient methodological rigor to withstand scrutiny. In pursuing this approach, we shifted the evaluation paradigm to a legal framework. In this framework, if the conclusion is that training adds value, the evidence needs to be persuasive, but not necessarily unequivocal. Although this is a lesser standard of proof than might be obtained through experimental methods, this evidentiary standard is sufficient to reach clear and reasonable conclusions about the value of a program.

In programs designed to enhance performance of people and organizations, defensible evidence can be obtained by establishing a chain of evidence showing the effect of a training program from the perspectives of multiple stakeholders at multiple points in time. The chain of evidence starts by establishing a clear understanding of expectations and an assessment of the return on those expectations (ROE). Value can be claimed when the ROE is high, and a high ROE is achieved when evidence shows that a program has achieved the three Cs: consensus, correspondence, and consistency.

Consensus—The First "C"

Stakeholder Involvement

An assessment of value starts with the question: "What outcomes do we expect from this program?" This sounds like a straightforward task—simply name the expectations. But whose expectations? Typically, training programs have multiple stakeholders who maintain different beliefs about program impact. Executives' concerns may reflect strategic issues (e.g., will this program help us respond to a shift in the market?), while those who build training programs will be concerned with tactical details (e.g., will a group activity best convey this content?). Given that multiple viewpoints exist and differ, defining value means populating expectations of impact with these multiple perspectives. For example, in

developing a training program these multiple perspectives might include content experts, project managers, instructional designers, and other stakeholders. Successfully measuring value requires integrating these diverse expectations into a single consensus view of what the program is supposed to achieve. Expectation setting requires three steps.

Step 1: Setting Expectations

First, in setting program expectations stakeholders articulate detailed and specific statements that describe the outcomes they expect from the program. Because different stakeholders hold different perspectives, and because we want to be inclusive in setting expectations, the list of outcomes tends to be numerous as well as diverse (see Figure 1).[i]

Step 2: Create the Key Messages About Training Impact

After the exhaustive list of outcomes is complete, stakeholders individually organize or sort the complete list of outcomes in Figure 1 into a smaller number of key concepts. Obviously, each person will approach the task differently. The key task here is to integrate these unique perspectives. A variety of group-consensus processes could be employed to achieve integration. However, we are often faced with the dilemma of including important perspectives from a geographically diverse stakeholder group. In this case, a face-to-face group process risks losing critical input. Our solution—the task of sorting, although it does not appear to be the case on the surface, provides rich quantitative information. By extracting this quantitative data, the data for all stakeholders can be integrated and clusters of outcomes can be identified (see Endnotes for a description of this analysis). The result is consensus on the key content areas that emerge from the detailed items. For example, based on input from multiple stakeholders, we determined that the detailed outcomes from Figure 1 could be aggregated into four key concepts (see Figure 2).[ii]

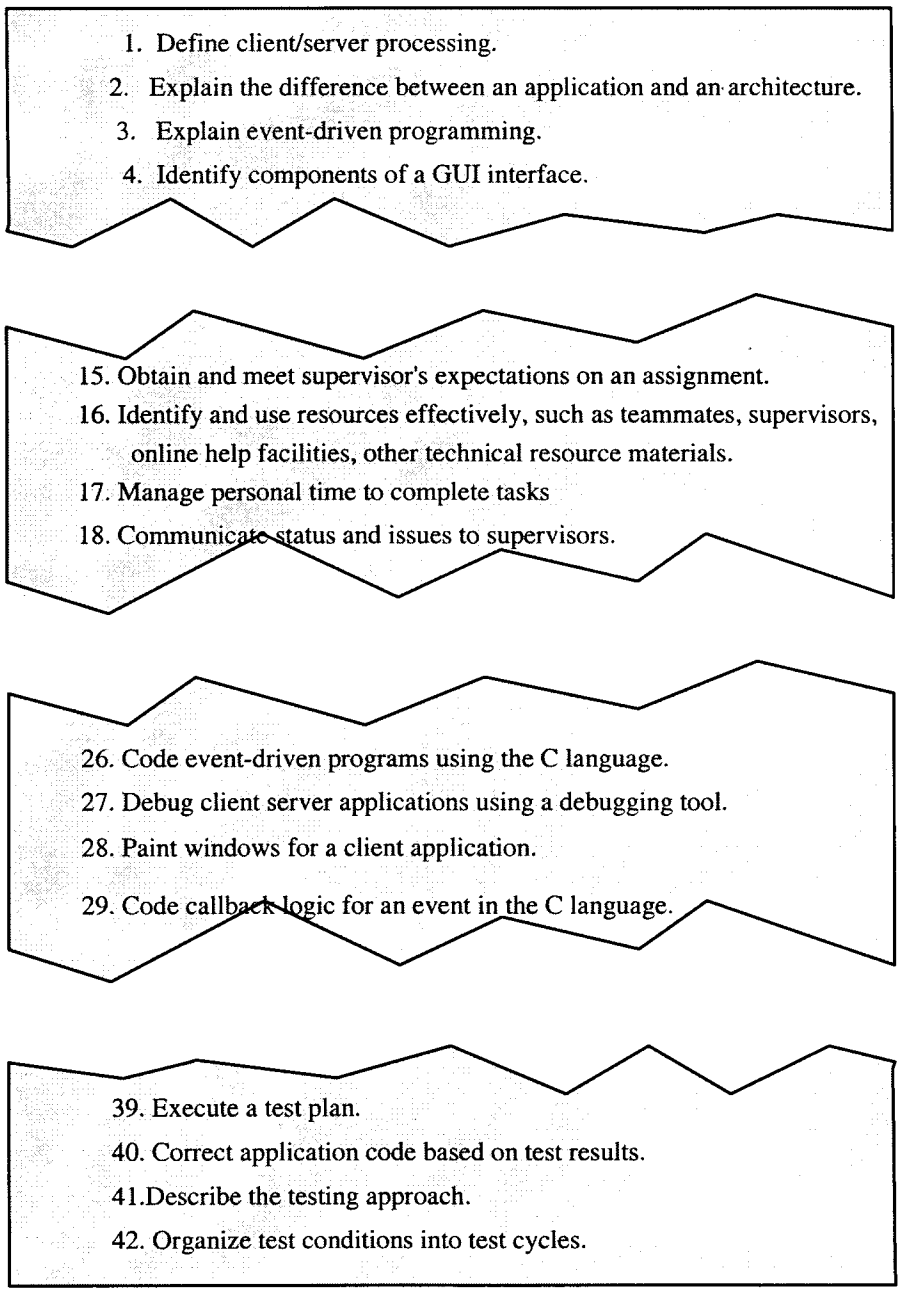

Figure 1. Create the Detailed List of Expected Outcomes with Input from Sponsors, Content Experts, and Instructional Developers.

Step 3: Create Expectations for Baseline Performance

Finally, stakeholders individually rate the extent to which the program can have an impact on each of the expected outcomes. That is, given the limited time for training, complexity of the task, and so forth, stakeholders are asked to consider the level of achievement they feel is reasonable for students to attain on each outcome. The assumption here is that students are not expected to achieve at the highest levels on all outcomes at the conclusion of the training program. Some outcomes are complex and will require work experience before the student is fully proficient. In that case, the training program ought to be held accountable to introduce and build basic proficiency. For other outcomes, training ought to quickly yield high proficiency, and consequently the training program ought to be accountable for building high proficiency by the conclusion of the training program. Creating baseline expectations means articulating the variation of intended impact among all of the outcomes.

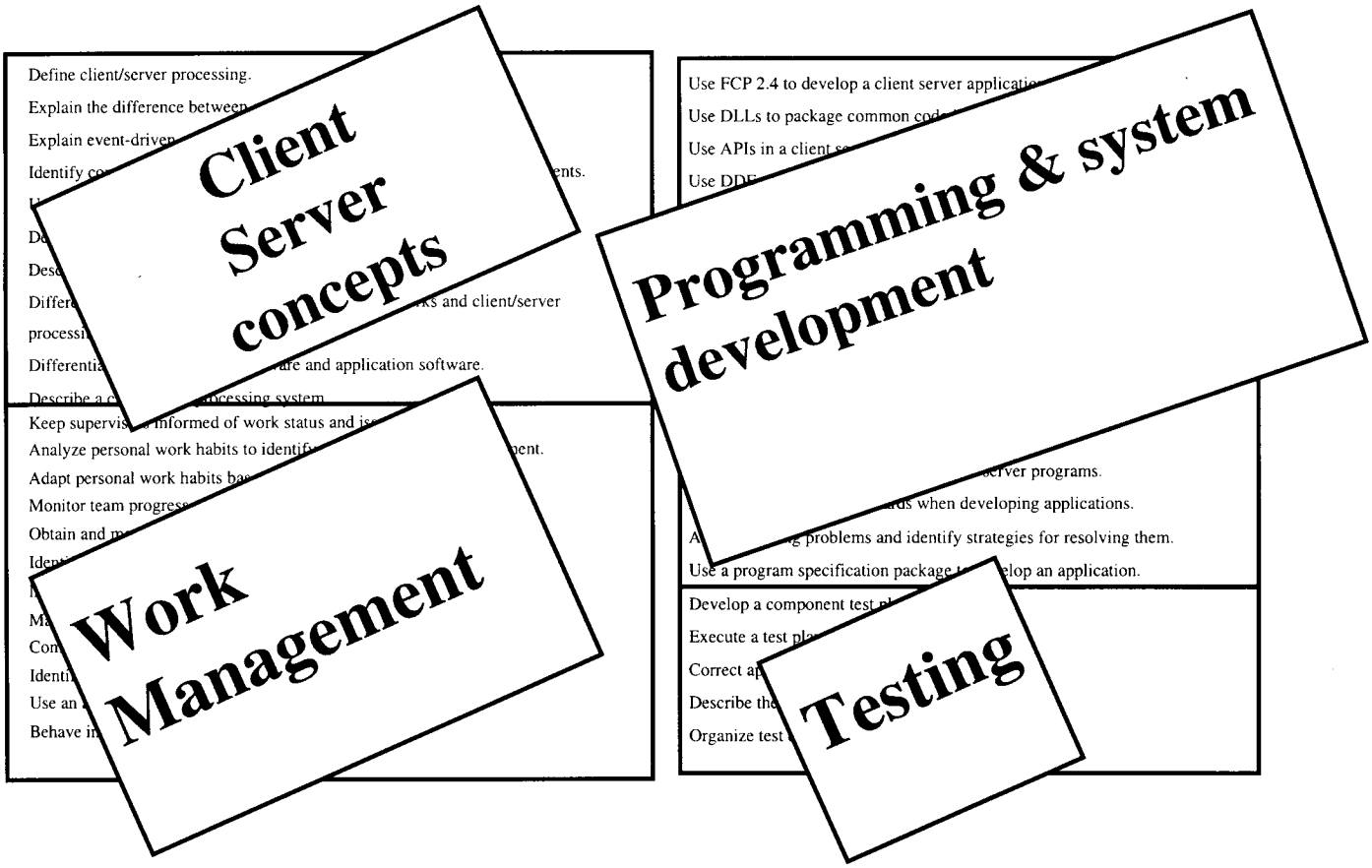

Figure 2. Organize the Detail into Key Messages About Training Impact.

Construct the Baseline Picture

The expected proficiency ratings provided by each stakeholder are averaged and displayed on a vertical number line (see Figure 3). This picture shows the consensus view of the expected impact of each major content area. In effect, this picture is a baseline of the expectations against which we can measure results.

Having set expectations for the program, the first of the three Cs (consensus) was accomplished. In this case, students were expected to quickly achieve a relatively high degree of competence on the "soft" skills (e.g., teamwork, personal management, client interaction, etc.). Because these typically are new employees who are new to the professional service environment, achieving high proficiency in these skills was important. Furthermore, the pattern of expectations provides a picture of the training program that highlights the fact that technical skills (e.g., programming) must be integrated with soft skills (e.g., work management) to successfully deliver value (e.g., a trained individual who can apply technical skills to provide client service). We were now in a position to evaluate the correspondence of outcomes with these expectations.

Correspondence—The Second "C"

Obtain the Student Perspective

Successful measurement of outcomes also requires integrating views between stakeholders who set expectations and those who assess the outcomes of training (e.g., students, supervisors, clients). Our definition of value asserts that the results or outcomes of training ought to "correspond" with the expectations for training. For example, if we ask students to judge the extent to which they have mastered the diverse set of skills underlying the four concepts in Figure 2, a match of the patterns should occur. Visually this means summarizing student data into concepts and arranging this data on a vertical number line in the same way Figure 3 was developed.

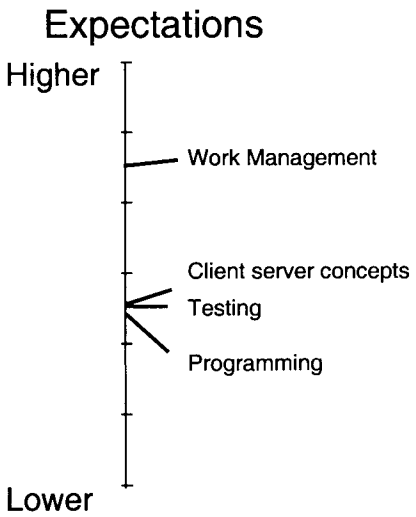

Figure 3. Portray the Consensus View on Expected Outcomes.

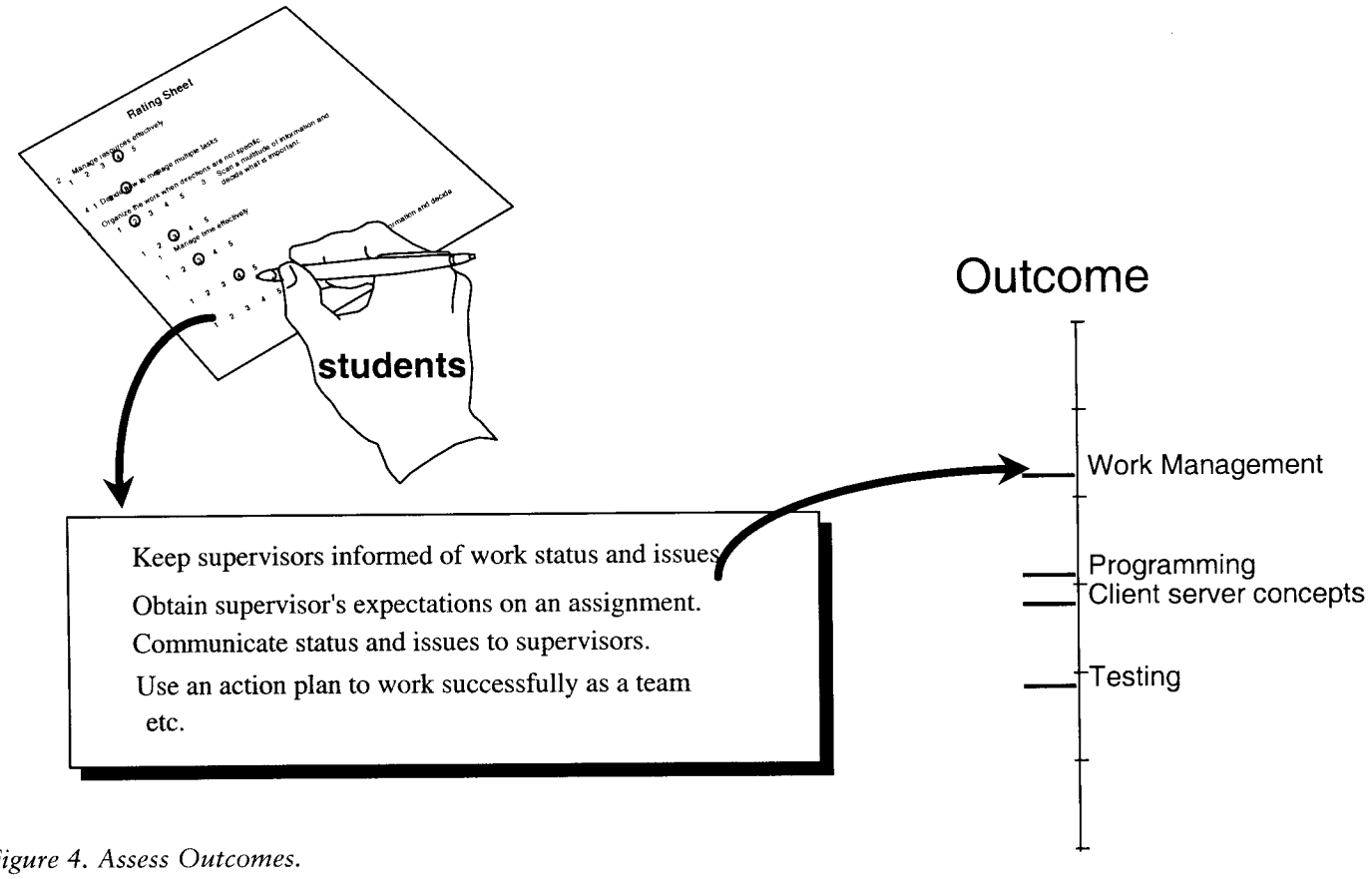

Figure 4. Assess Outcomes.

Compare the Student Perspective Against Expected Impact

To visualize correspondence, join Figure 3 (the stakeholders' expectations for impact) on the left and Figure 4 (students' self-assessment of proficiency) on the right to produce Figure 5. The alignment of concepts in Figure 5 shows that, for this program, the expectations correspond to student proficiency at the program's conclusion. In other words, the expectations and outcomes align, or the patterns on the left and right match. If perfect, this picture would resemble a ladder. As a program departs from perfection, lines tend to slope, and in the extreme, lines will cross.

In this example, the news from the learning event is positive; alignment is high. On the other hand, while correspondence between expectations and outcomes at the learning event is certainly good news, it is not sufficient. Rather than a single point, a chain of evidence is necessary.

Triangulate Using Multiple Perspectives

Our approach has been to collect evidence from a variety of perspectives: the perspective of students at the learning event, former students who now have the experience of applying skills in a work environment, and supervisors who manage the work of former students. Each of these points of inquiry is based on the tactic of obtaining input on the detailed school outcomes (see Figure 1) and directly comparing this data to baseline expectations (see Figure 3).

Consistency—The Third "C"

Training or other programs tend to be conducted multiple times. One session should not be high on testing and another session low on testing. Although students at both sessions may express satisfaction with the event, the outcomes are not consistent. Two things are apparent. First, the training experience needs to be the same across the organization. If the different students learn different things, the organization is not building a common knowledge base. Second, the investment of resources (e.g., personnel to support the training event, training methods, and student and faculty materials) in the delivery of programs ought to ensure consistent outcomes. In other words, there ought to be a return on the investment in the delivery of the program. Expanding the idea of Figure 5 over time (see Figure 6) shows, in this case, that outcomes at the learning events are consistent. The outcomes from one session are aligned with the outcomes from the next session of the same program.

Although students and faculty may differ over time, the outcomes remain stable. Two conclusions are apparent: students in different sessions have similar educational experiences, and the investment in resources to support the program are ensuring consistency.

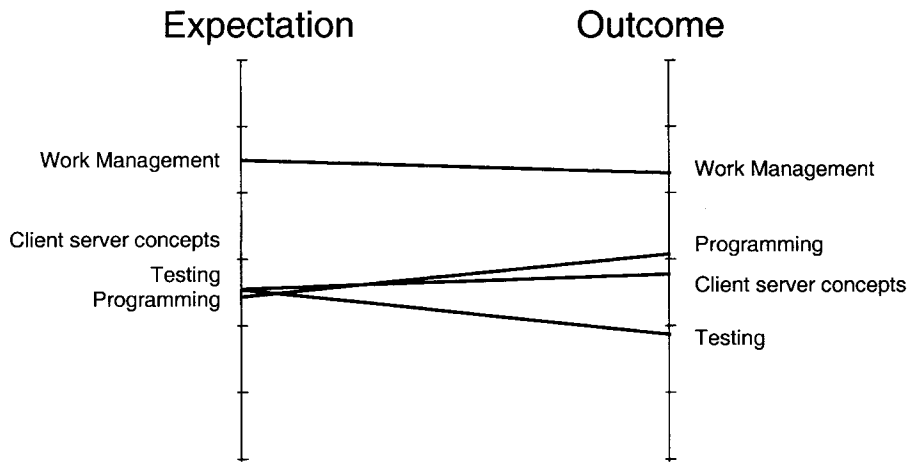

Figure 5. Determine the Correspondence Between Expectations and Outcomes.

Conclusions

Our intent has been to expand the interpretation of "return," the methods of quantifying "return," and to illustrate this process. Although more data points exist for the program used in this illustration, all points show basically the same pattern for this program—a nearly parallel set of lines illustrating alignment of outcomes and expectations. By getting to parallel, this program demonstrates value.

The chain of evidence is such that we can assert this program meets expectations and produce a preponderance of evidence to support that assertion. Specifically, the data points align well or correspond to program expectations—student learning at educational events, former student views on school effectiveness, supervisor views on the competence of program graduates, and supervisor views on the importance of school content in the work environment.

What about cases in which the alignment is not parallel? The pattern-matching process highlights areas not in alignment and quickly focuses discussions on program improvement. For example, if a content area is out of alignment, the data from the detailed items that make up the concept can be analyzed. This is analogous to turning up the power on a microscope to examine finer details. For instance, if testing was out of alignment, we could analyze the data for each of the items that make up testing. Our analysis would pinpoint which items are low and focus discussion on what it takes to effect a change in these items. In other words, by turning up the power of the evaluation microscope, we can look at finer detail, which in this case provides diagnostic data to improve the program.

In addition to broadening the definition and methods of evaluation, our intent has been to raise the issue of the return expected from the investment in evaluation efforts. Evaluation methods need to be adaptable to the complexities of organizations. Specifically, creating evidence of value over multiple programs means evaluation methods need to provide compelling evidence with minimal tailoring. The ROE approach provides an additional tool for responding to questions of value. It also offers a process that accommodates the uniqueness of the expectations for individual programs, that is, a rigorous and repeatable process that accommodates variable content.

Endnotes

i The complete list consists of 42 statements; due to the proprietary nature of the content, a representative sample of outcome statements are presented here.

ii Soliciting and integrating input from multiple key stakeholders with varying perspectives creates a challenge. Specifically, how can individual beliefs about the training program be efficiently combined into a single coherent group perspective—by applying the statistical tools of multidimensional scaling (MDS) and cluster analysis (CA). When individuals sort outcomes, they make explicit their

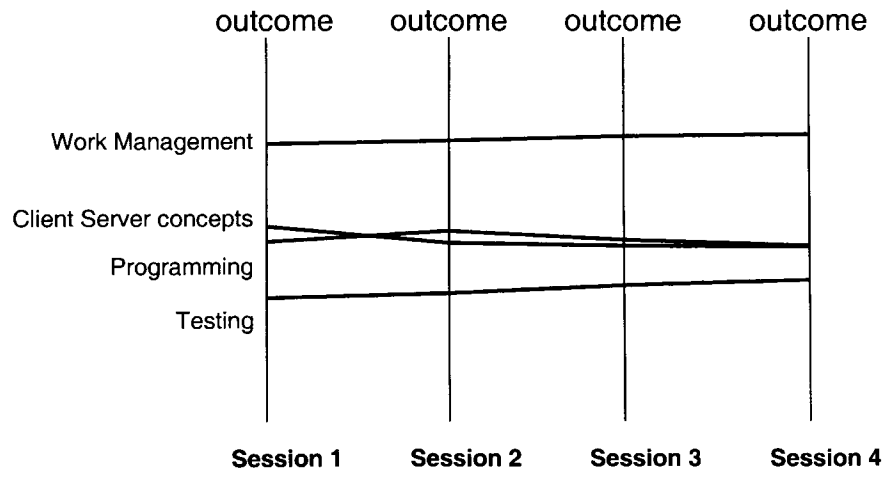

Figure 6. Assess the Consistency of Outcomes over Time.

beliefs about the organization of those outcomes. That is, outcomes that are similar get placed together into the same groups. When multiple people sort outcomes, in effect creating many different groupings of the outcomes, MDS can determine the degree of similarity of each outcome with every other outcome for the entire group of stakeholders. When CA is applied to this data, groupings or clusters of outcomes can be identified—for example, the training program in this paper was best described with four themes (work management, programming, client server, testing). However, reaching the conclusion that four themes is appropriate is not simply an analytical process. This type of analysis does not supply one single correct answer; rather this analysis provides multiple solutions. For example, if we thought five themes might best express the intent of the program, the statistical algorithms can show us where to cut the four groups to make five. Although four themes best represented the training program in the case example, the final choice was reached after much discussion and consideration of other solutions (i.e., three, five, and six clusters) by the decisionmakers.

International Society for
Performance Improvement

October 1998
Volume 37 / Number 8

performance improvement