From the Beginning

These two articles can help your training program get off to the kind of start that will make it all worthwhile.

Using Delphi for Curriculum Development

Contributed by Miles O. Weaver, a consultant based in Fremont, Ohio.

One of the most frustrating obstacles presenters of "state-of-the-art" subjects encounter is being sure that the content of their presentations is, in fact, timely and relevant. This is particularly true for educators and trainers whose clients are business and industry.

Training professionals can use the Delphi method to obtain a consensus from a panel of experts within a given field. After methodically collecting and carefully summarizing the opinions of all panel members, the trainer submits the summary to the panel members again and asks those experts to rethink their opinions. The trainer repeats the process until the experts reach a consensus or until it is apparent that no consensus is possible on the given issue.

The Delphi method offers the following desirable features:

- You obtain opinions from a *selected* group of experts.
- Repeatedly polling the jury in the light of the latest group consensus leads to an interaction between the individual and the group.
- Since you do the polling with the individual separated from the rest of the jury, you minimize dictatorship by a vocal minority.
- Repeated polling gives the individual participant a sense of involvement and ownership in the program.

A typical Delphi study

In 1985 I used the Delphi method to generate guidelines for a statistical process control (SPC) training program. I conducted a survey concerning the desired content of SPC training materials and solicited responses to questionnaires from members of local chapters of the American Society for Quality Control (ASQC) and from other practicing quality control professionals. One of those questionnaires appears on page 19. I had t ıker the topics included in the questionn aire principally from the most current A SQC

Figure 1—Summary of relative responses

5 Extremely Important

- 4.8 Measurement
  - Need for quality
  - Process capability
  - Process control
- 4.7 Control charts
  - Control limits
  - Decision making
  - Machine capability
  - Measurement error
  - Standard diviation
  - Statistical control
- 4.6 Accuracy
  - Average
  - Gauge Capability
  - Histogram
  - Measures of central tendency
  - Normal distribution
  - Pareto analysis
  - Quality
  - Quality control
  - Specifications
  - Statistics
- 4.5 Measures of dispersion
  - Sampling
  - Variables
  - Variation
- 4.3 Distribution of data

Figure 2—Summary of relative learning difficulty responses

3 High Learning Difficulty

- 2.9 Weibull distribution
- 2.8 ASQC certification
- 2.7 Analysis of variance
  - Design of experiments
  - Inferential statistics
  - Regression
- 2.6 Chi-square analysis
  - Coefficient variation
  - Hypergeometric probability distribution
  - Hypothesis testing
  - Nonparametric statistics
  - Statistics
  - Tchebycheff's inequality
- 2.5 Coefficient of determination
- 2.4 Alpha error
  - Alpha risk
  - Coefficient of correlation
  - Exponential distribution
  - Machine capability
  - Normality, tests for
  - Poisson distribution
  - Probability
  - Process capability
  - Reliability
  - Report writing
  - Standard deviation
- 2.3 Beta error
  - Beta risk
  - Proability distribution
  - Process control
  - Statistical control

certification brochure and from the SPC textbooks and references used in the Quality Control Technology program at Terra Technical College in Fremont, Ohio.

I asked the jury to identify each topic's relative importance and relative learning difficulty. I designed the importance rating by using a five-point scale from 1 (no importance) to 5 (extreme importance) and the learning difficulty rating by using a three-point scale from 1 (low difficulty) to 3 (high difficulty). I then analyzed the results of the initial survey, summarized them, and resubmitted them to the respondents. I asked the respondents to try to modify their original responses based on the overall summary of the group's individual responses.

I ordered from highest to lowest value the topics and their mean rating values for both relative importance and relative learning difficulty. Figures 1 and 2 show the relative importance and relative learning difficulty rankings. I also organized all

Initial Delphi questionnaire

Name: _____

Title: _____

Company: _____

Address: _____

_____

Telephone No.: _____

1. In your opinion, is there a need for an SPC text specifically designed for use by people without an engineering or science education? *Yes/No* (Circle)
2. Please rate each of the following topics on a five-point scale as to their importance and on a three-point scale as to learning difficulty in an SPC textbook to be used for a full spectrum of personnel in a manufacturing plant training situation. These people will *not* be trained statisticians as a result of this training but will be expected to make process control assessments on a routine basis on the production line, in a QC laboratory, etc.

Topic	Importance					Learning Difficulty		
	None	Marginal	Moderate	Very	Extreme	Low	Medium	High
Acceptance sampling	1	2	3	4	5	1	2	3
Accuracy	1	2	3	4	5	1	2	3
Alpha error	1	2	3	4	5	1	2	3
Alpha risk	1	2	3	4	5	1	2	3
Analysis of variance	1	2	3	4	5	1	2	3
Arithmetic mean	1	2	3	4	5	1	2	3
Array	1	2	3	4	5	1	2	3
ASQC certification	1	2	3	4	5	1	2	3
ASQC membership	1	2	3	4	5	1	2	3
Attributes	1	2	3	4	5	1	2	3
Average	1	2	3	4	5	1	2	3
Beta error	1	2	3	4	5	1	2	3
Beta risk	1	2	3	4	5	1	2	3
Bias	1	2	3	4	5	1	2	3
Binomial distribution	1	2	3	4	5	1	2	3
Business goals	1	2	3	4	5	1	2	3
c Charts	1	2	3	4	5	1	2	3
Capability index	1	2	3	4	5	1	2	3
Capability ratio	1	2	3	4	5	1	2	3

Figure 3—Typical organization of subject material

	Topic	Learning Importance	Difficulty
I.	**Introduction**		
	Quality	4.6	1.8
	Quality control	4.6	1.8
	Need for quality	4.8	1.9
	Business goals	4.1	1.5
	Organizational relationships	3.6	1.8
	Deming's 14 points	3.6	1.5
	History of quality control	3.2	1.3
	ASQC membership	2.5	1.2
	ASQC certification	2.7	2.8
II.	**Calculation Methods**		
	Basic computer methods	3.3	1.8
	Coefficient of correlation	2.8	2.4
	Coefficient of determination	2.6	2.5
	Coefficient of variation	2.9	2.6
	Computer calculations	3.8	2.1
	Exponents	2.9	2.1
	Hand calculator operations	4.5	1.8
	Product quality index	3.1	1.9
	Scientific notation	3.0	2.2
	Signed numbers	3.1	1.9
	Transformations of data	3.3	2.2
III.	**Metrology**	3.8	2.3
	Measurement	4.8	2.1
	Fundamental units	3.5	1.5
	Standards	4.3	1.9
	Error	4.1	1.9
	Measurement error	4.7	2.2
	Accuracy	4.6	1.8

the topics in accordance with the sequence and categories I had used for industrial SPC training sessions in the past. Figure 3 shows a typical page from the resultant summary, which also contains the mean rating values for relative importance and relative learning difficulty for each topic included.

The results you obtain in a Delphi study represent a firm, locally valid basis on which you can build a relevant curriculum. The results also can serve as guidelines for generating a detailed text and lecture notes you can use in training sessions. The relative importance values suggest the *emphases* you should give to the various topics. Similarly, the relative learning difficulty values indicate the *amount of time* you should allocate for a given topic.

A starting point

The Delphi method is a promising technique for eliciting the recommendations of a panel or jury of experts concerning the content of a meaningful and timely course of study in formal educational or "stylized" training programs. The results of such a study provide a reliable and valid starting point for building a training program to meet the needs of prospective client firms represented by the participants in the study.

This method offers another advantage, as well: it enlarges the support base for the program within the social and industrial communities of the study participants.

Integrating Evaluation, Design, and Implementation

Contributed by Susan M. Connolly, president of Connolly Associates in Webster, New York.

If it is to have any real impact on the quality of training itself, training evaluation should be closely linked to the processes of design and implementation. Trainers can achieve this integration—but only if they involve curriculum designers and instructors in the evaluation process.

Too often designers and instructors are merely the recipients of evaluation results, often causing them to be defensive because they haven't been involved in the evaluation process. If these people can play a more active role in determining how the evaluation is conducted, the results could help designers and instructors achieve continuous quality improvements.

Trainers can link evaluation to curriculum design at four distinct points in the design process: development testing, postpilot assessments, posttraining evaluation, and follow-up evaluation. Figure 1 illustrates how to accomplish this integration.

Development testing

Development testing involves trying out pieces of the program, independently or in sequence, to see how well they work before they are tested in a classroom setting. Trainers should conduct these tests with representatives of the intended audience during the formative stages of program design. At this point the curriculum designer has defined the training objectives and is in the process of identifying content, determining instructional strategies, and preparing course materials.

The procedures for development testing vary according to the purpose of the test and the nature of the material to be tested. Trainers who are conducting development tests for the purpose of assessing overall program flow and organization may need to schedule pilot tests to run through everything in sequence.

More often, however, development testing is a less formal procedure. The designer may want to see how a new group exercise or skill practice works. He or she can accomplish this by gathering together a few potential participants and trying the material. The designer should inform participants of the session's purpose beforehand. During the session, the designer should administer the exercise just as it would be in an actual classroom situation. While observing how the materials actually work out, the designer can also solicit participant reactions to the materials by debriefing after the exercise is completed.

Henrietta Komras has suggested that trainers should use the development test to identify problems in four categories: language clarity, difficulty, design problems, and interest level. The development test should also serve as a "reality check": Do the materials have credibility in terms of "real world" simulation?

Development testing is beneficial for the design process in the following ways.

- It helps the designer identify problems that he or she can resolve prior to the pilot phase of the program.
- It helps to establish time requirements for various exercises.
- It allows the designer to approach the pilot with increased confidence.

Postpilot assessments

Trainers should conduct this second phase of the evaluation process during and immediately following one or a series of pilot programs. They should collect postpilot data from program participants, instructors, curriculum designers, and subject matter experts. Trainers must carefully select pilot participants who represent future audiences, and they should tell participants that they are taking part in a pilot effort and that their candid feedback will contribute to program improvements.

Trainers can gather information from participants in a number of ways:

- by asking them to complete end-of-course questionnaires that solicit closed and open-ended responses to various aspects of the program;
- through brief end-of-module discussions interspersed in the program to assess immediate reaction to specific materials;
- by less formal discussions with participants during breaks.

On the other hand, trainers should establish a systematic procedure for collecting and documenting feedback from designers and subject matter experts who observe the pilot programs. Trainers can provide these observers with observation checklists that require them to focus and comment on specific aspects of the program. Lynne Tyson and Herman Birnbrauer have suggested using a "Learning Conditions Checklist," which asks observers to respond "ok" or "not ok" to a series of specific items under the following categories:

- clarity of objectives and appropriateness of content;
- trainees' demonstrated performance;
- trainer's demonstrated performance;
- organization and sequence;
- time allocations;
- training strategies, materials, and equipment;
- facilities and logistics;
- transportation;
- administration and management.

It is useful to hold a debriefing session the day after the pilot program. The program instructor and all observers should attend this session. This is the time for key program contributors to discuss their observations and, perhaps, reach consensus on needed changes. This session will be more productive if all attendees have documented their observations in some systematic manner. The major benefit of postpilot assessments is in identifying needed program revisions at a time when it is still relatively easy to make them.

Posttraining evaluation

Once pilot materials are revised and finalized, the training program is usually implemented on an ongoing basis. At this point the evaluation takes on more summative characteristics. Now is the time to look at the product's effectiveness.

The third and fourth phases of the evaluation process should assess the four areas Donald Kirkpatrick has described: reactions, learning, behavior, and results. Phase three is an immediate posttraining measurement and includes two components: reactions and perceptions, and learning acquisition.

Reactions and perceptions. This measurement consists of self-report feedback from the training participants. Trainers should collect the data via an end-of-course questionnaire that includes both closed and open-ended questions. Trainers can used this questionnaire to assess participants' reactions to many aspects of the program including

- importance and relevance of content;
- value of exercises;
- pace and length of the program;
- quality of materials;
- quality of instruction.

Trainers also can use this end-of-course instrument to collect participants' perceptions relative to the on-the-job value of the training.

Due to its' subjective nature, the end-of-course self-report form is usually considered to be the weakest form of evaluation. Nevertheless, it is essential as an initial measurement. After all, if participants react negatively to the training and perceive that it has no value, it is unlikely that they will learn and apply new skills. And if trainers keep this questionnaire short and simple, and if it lends itself to computerized analysis, this instrument becomes an excellent monitoring device that trainers can use for the life of the program.

Learning acquisition. Trainers can use objective measures during and immediately after training to determine whether participants have actually acquired the intended skills, knowledge, and attitudes. If trainers want to be able to attribute gains to the training, they should compare postcourse measures to precourse measures. It isn't necessary to implement learning acquisition measures for the life of the program. Trainers can use them for a period of time to validate the learning and then can discontinue them.

Trainers can collect data via paper-and-pencil tests, systematically observing classroom performance, and examining materials produced during the training. It is important for designers to integrate creatively these collection tools into the program design so that they don't have the appearance of tests.

If integrated effectively, these measures can both evaluate and contribute to learning. For example, a training program in public speaking may require participants to deliver several short presentations to the class. These presentations could be videotaped and then assessed against specific learning criteria. In this way, trainees learn by viewing their performance and trainers can document learning and measure it objectively.

It is true that skill acquisition is easier to measure than knowledge acquisition, but trainees can demonstrate that they understand concepts or procedures by writing them down. Designers can integrate these paper-and-pencil measures into the training as end-of-module or end-of-session reviews.

Follow-up evaluation

Evaluation results from the third phase may indicate that participants liked the program, they perceived that it has value, and they acquired the intended knowledge, skills, and attitudes. Now, in phase four, it is important to answer several questions: Are participants retaining the learning? Are they using it on the job? Does the learning make any difference?

Phase four of the evaluation process should take place three to 12 months after the training is completed. It includes the three measurement components: learning retention, on-the-job application, and organizational impact.

Learning retention. The purpose of this measurement is to determine whether trainees have retained the knowledge, skills, and attitudes learned in the classroom. Learning retention should not be confused with learning application. It is possible for a trainee to retain learning and still not apply it on the job. Furthermore, lack of retention is only one of the

possible reasons trainees fail to apply skills.

Trainers can apply the same methods they use to measure learning acquisition to measure learning retention. The major difference here is that trainers should assess learning retention at least three months after the training is finished. They can then integrate these measurements into the training programs by administering the instruments during a follow-up refresher session.

On-the-job application. The ultimate test of training effectiveness is whether trainees use the skills on the job. This is what Kirkpatrick described as measurement of behavioral change. Trainers can gather this kind of data via reported or observable measures. Some combination of both methods is usually the best approach.

Reported measures can include questionnaire or telephone surveys and critical incident interviews. Trainees, their direct reports, and their supervisors are all sources of data. Trainers must skillfully design the questions to illicit specific examples of behaviors that respondents perceive to be a direct result of the training. Some people view reports from the trainees themselves suspiciously, but I've conducted research that has shown that trainee self-reports are at least as valid as those of their subordinates and managers.

Observable measures can include examining records or products and conducting on-site observations. For example, trainers could assess the effectiveness of a performance appraisal workshop by examining the written performance appraisal forms trainees produce after returning to the job. Or, trainers could assess the application of skills learned in a CAD-CAM program by examining the designs the trainees produce back on the job.

Keep in mind, however, that trainers must base any examination of records or products on a predetermined set of criteria and conducted in a systematic manner. On-site observations are an effective, but seldom-used, means of gathering data. Chris Stark has proposed a technique called the "Sit-In" that tends to reduce the potentially awkward nature of on-the-job observation. This method suggests that the trainer can play an important role in reinforcing the use of newly acquired skills. The sit-in combines training evaluation with on-the-job coaching and requires

careful contracting with the trainee and his or her management to establish clear initial guidelines.

Trainers measuring on-the-job skill application should attempt to go beyond merely establishing the degree to which trainees apply learning. It is also important to identify the environmental factors that contribute to or interfere with application. The results may indicate that some of these factors are beyond training's control.

Organizational impact. This is the most difficult aspect of training evaluation. It establishes the value of the training to the organization. Assessing organizational impact requires identifying, monitoring, and measuring indirect environmental factors. For example, trainers could measure the organizational impact of a program on safety procedures by tracking the number of accidents. But in order to attribute results to the training, trainers must be able to compare pre- and posttraining measures.

It isn't always possible to link training results to highly measurable factors such as rejection rates, profits, turnover, and attendance. But with the current emphasis on quality and efficiency, there will be more and more emphasis on the need to establish the organizational value of training.

The process outlined above describes a comprehensive training evaluation approach. The demands of a business environment seldom allow trainers to use all four phases, but this isn't an excuse for doing nothing because each phase in the process has value. The curriculum designer needs to decide which aspects are most useful and incorporate them into the overall training design and implementation plan.