DOCUMENT RESUME

ED 069 736 TM 002 191

AUTHOR Cogan, Eugene A.; Lyons, J. Daniel
TITLE Frameworks for Measurement and Quality Control.
INSTITUTION Human Resources Research Organization, Alexandria, Va.
REPORT NO HumRRO-PP-16-72
PUB DATE Jul 72
NOTE 16p.; Presentations at New York University First Annual Training in Business and Industry Conference, New York, March 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Conferences; *Job Analysis; *Measurement; *Performance; *Personnel Selection; *Quality Control; Recruitment; Training Objectives; Work Simplification

ABSTRACT

Two presentations are offered on frameworks for measurement and quality control in business and industry. The first paper states that in making a job performance evaluation, anything that can be specifically defined can be measured. However, to develop a testing program that is both useful and cost-effective, it must be known who will make what decision, using the obtained measurements. Analysis and interpretation of the particular purpose and setting are needed. Feedback data show how improved decisions can produce dollar gains far beyond the cost of developing and employing measurement. In the second paper, the essential elements of a quality control system are illustrated, including: (1) training objectives or performance requirements, (2) proficiency and diagnostic measures, (3) data reductions and analysis, (4) procedures for decision and corrective action, (5) communication procedures, and (6) managerial support. It is shown that training goals must be defined in terms of measurable on-the-job performance. (Author/LH)

Professional
Paper
16-72
HumRRO-PP-16-72

HumRRO

TM

Frameworks for Measurement and Quality Control

Eugene A. Cogan and J. Daniel Lyons

Presentations at
New York University
First National Annual
Training in Business and Industry Conference
New York City
March 1972

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street • Alexandria, Virginia 22314

July 1972

The Human Resources Research Organization (HumRRO) is a nonprofit corporation established in 1969 to conduct research in the field of training and education. It is a continuation of The George Washington University Human Resources Research Office. HumRRO's general purpose is to improve human performance, particularly in organizational settings, through behavioral and social science research, development and consultation.

Published
July 1972
by
HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street
Alexandria, Virginia 22314

BIBLIOGRAPHIC DATA SHEET	1. Report No. HumRRO-PP-16-72	2.	3. Recipient's Accession No.
4. Title and Subtitle FRAMEWORKS FOR MEASUREMENT AND QUALITY CONTROL			5. Report Date July 1972 6.
7. Author(s) Eugene A. Cogan and J. Daniel Lyons			8. Performing Organization Rept. No. PP-16-72
9. Performing Organization Name and Address Human Resources Research Organization (HumRRO) 300 North Washington Street Alexandria, Virginia 22314			10. Project/Task/Work Unit No. 11. Contract/Grant No.
12. Sponsoring Organization Name and Address			13. Type of Report & Period Covered Professional Paper 14.

15. Supplementary Notes
Two presentations at New York University First National Annual Training in Business and Industry Conference, New York City, March 1972.

16. Abstracts The author of the first paper states that, in making a job performance evaluation, anything that can be specifically defined can be measured. However, to develop a testing program that is both useful and cost-effective, it must be known who will make what decisions, using the obtained measurements. Analysis and interpretation of the particular pupose and setting are needed. Feedback data show how improved decisions can produce dollar gains far beyond the cost of developing and employing measurement. In the second paper, the essential elements of a quality control system are illustrated, including (a) training objectives or performance requirements, (b) proficiency and diagnostic measures, (c) date reduction and analysis, (d) procedures for decision and corrective action, (e) communication procedures, and (f) managerial support. It is shown that training goals must be defined in terms of measurable on-the-job performance.

17. Key Words and Document Analysis. 17a. Descriptors
*Performance evaluation
*Performance standards
*Performance tests
*Measurement
*Test construction
*Validity
*Quality Control
*Job analysis
*Personnel selection
*Work measurement
*Training systems

17b. Identifiers/Open-Ended Terms
Quality control systems
Instructional systems
Accountability

17c. COSATI Field/Group 05/09

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Distribution of this document is unlimited.	UNCLASSIFIED	15
	20. Security Class (This Page) UNCLASSIFIED	22. Price

FORM NTIS-35 (10-70)
USCOMM-DC 40329-P71

Prefatory Note

These papers were presented at the First National Annual Training in Business and Industry Conference of New York University, held in New York City in March 1972. The first paper, " 'If It Exists, It Can Be Measured'—But How?" was prepared by Dr. Eugene A. Cogan, who is Director for Research Design and Reporting in the Executive Office of the Human Resources Research Organization (HumRRO) in Alexandria, Virginia. The second paper, "Measuring Effectiveness: Quality Control of Training," was prepared by Dr. J. Daniel Lyons, who is Director of HumRRO Division No. 1 (System Operations), also located in Alexandria, Virginia.

"IF IT EXISTS, IT CAN BE MEASURED"—BUT HOW?

Eugene A. Cogan

Psychologists—including those especially interested in measurement—have been, and continue to be, plagued by elusive and fragile concepts. Many concepts have their origin in the individual and cultural experiences all people share. For example, we all have the feeling that we know some people who seem "smarter" than others over and above differences in their schooling or other educational experience; this feeling has led to the concept of "intelligence" and to attempts to define, understand, and measure intelligence. Our shared experiences have led us to feel that some people are better employees than others; this feeling has led to attempts to define, understand, and measure "goodness as an employee." Attempts to cope with "goodness as an employee" have been equally as frustrating to employers and to psychologists as have been attempts to make sense of "what is intelligence all about."

The main stem of the title of my paper—"If it exists, it can be measured"—is a free translation of a classic statement by Edward Thorndike who was trying to counter the pursuit of poorly defined pseudo-concepts that bordered on being personal illusions. For us, Thorndike's message is: "Until you can define what you are interested in well enough so that you can figure out how it can be measured, it can mean anything and, therefore, it means nothing."

The challenge of Thorndike's proposition to theoretical psychology has no easy answer because theoretical psychology *is* concerned with generally important abstractions regarding human behavior. There is an understandable reluctance to fix on formal definitions for concepts because useful definitions must be restrictive and omit things; theoretical psychologists are reluctant to risk throwing out a baby with the bath water.

However, for practical, applied measurement the implications of Thorndike's doctrine are very useful. In a practical setting, Thorndike's edict translates to: "Of course you can measure it, after you have defined what it is." The main purpose of my presentation will be to deal with how to go about defining "it" so that you can proceed to measurement, and then how to evaluate the measurement.

In any practical setting, there are many situation-specific features and these provide a key to measurement. The trick to translating an impression into a measurable something consists of using the situation to define what measurement is needed.

Purpose of Measurement

Foremost for defining measurement is "why." In selecting or devising a measurement, it is essential to decide or determine the purpose of the measurement. In industry, the purpose translates to decisions that management or personnel people must make. Who will decide what with the aid of measurement information?

It is not enough to stop analysis of purpose at the broad levels of selection, assignment, promotion, training evaluation, or personnel evaluation. Each of these includes so many variants that depend on *particular* purposes that the category is the beginning, not the end, of analysis. If concern is with selection, the proper measurement depends on whether selection is for training or for direct job assignment, whether concern is solely for competence in an entry job or also with potential for advancement,

whether the work setting is closely supervised or relies on self-supervision, whether the work setting requires team work or individual work, and so on.

Even what seems to be a specialized and highly specific purpose like quality control of training, as is shown in Dr. Lyons' paper, involves at least four distinctive purposes and each of these has its own distinct definition and measurement.

What is Measured

In a particular setting, with purpose established in terms of the particular decisions that are to be made, the second element in defining the measurement concerns what is to be measured. Much of the definition of "what" will already have been established in careful definition of purpose. That is, if the purpose concerns selection for a training program preceding assignment to a job, the "what" should not contain very many, if any, direct indications of job knowledges and skills, but rather should deal with ability to learn these knowledges and skills. On the other hand, if selection is for direct assignment to job duties, it is whether these have been previously learned that is pertinent.

The matter of what is to be measured has been, by far, the subject of most concern and debate in industry and among measurement specialists. Primarily, this is because dollars and time for measurement, cost elements that are very sensitive in industry, are heavily dependent on what is measured. For example, considering job performance evaluation, the best theoretical measure is unobtrusive, scientific observation and careful measurement of behavior, over a long period of time, in the actual job setting. While such measurement is technologically possible, it would be so prohibitively expensive that less costly alternatives are always being sought and, typically, used. However, these less costly methods do not measure the same thing.

Usually considered closest to scientific observation in the natural setting is a job sample test. Even assuming that sampling of the job performances is well done, job sample simulation is not the same as job observation because important contextual and personal elements cannot be simulated. That is, a test environment creates test performance for the individual. He may try much harder than he does in the natural setting, or he may be immobilized by test anxiety.

Less costly—and hence more common than job sample simulation tests—are analytic tests of job performance elements. Such tests measure component skills and knowledges underlying job performance. We are all familiar with such analytic tests as they apply to selecting a secretary. For a candidate secretary, one might use a typing test, a dictation test, and a spelling test. While such tests can provide assurance that necessary individual job skills are within the candidate's repertoire, they do *not* assure the person can fit the skills and knowledges together effectively in a job setting, or that the person can or will do the many other tasks required on the job.

For still less cost than analytic tests, there are indirect tests of capabilities, usually paper-and-pencil tests dealing with incidental information about the job.

The simplest of the indirect tests are specialized vocabulary tests. For example, a good secretary is likely to know what "platen" means, and what a number four pencil is, and what the term "stay-back file" means. Since none of these three items of information is intrinsically of consequence in doing a good job as a secretary, they constitute *indirect* measures.

Use of indirect measures must be approached with great caution and checked empirically against more direct measures. This is because possessing such information may not come from job competence—witness the fact that I know the meaning of the three terms, but I have no secretarial competence whatsoever.

Most common of all as a measurement of job performance in industry is the rating scale. The reasons are that, first, it is the least expensive measure and, second, it seems to make sense to go to the day-to-day observer of job performance who has "seen job

performance with his own eyes over a long period of time." Despite the sensibility and low cost of rating scales, they don't do what most people think they do. Rating scales—regardless of what the rater is asked to check—provide a measure of an overall "Joe is OK by me," rather than how well Joe can perform elements in his job. I do not at all intend to deprecate the value of personnel decisions based on "Joe is OK by me"; I wish, however, to emphasize that what is being measured in that fashion differs from what is measured by a performance test even if the terms used are similar.

There are differences in what is being measured for all the categories named: natural observation, job sample tests, analytic tests, indirect tests, and rating scales. Treating them as alternate techniques to measure the same thing can be severely misleading. It is traditional to consider these measurements as alternatives, differing in technique but not in what is being measured. This inaccurate assumption of equivalence is possible only because not enough—and not precise enough—analyses have been performed to define purposes of measurement and what is to be measured.

Effectiveness of Measurement

I will now turn to effectiveness or—in psychometric terms—validity, as it applies to the consideration of measurement.

I began this paper with the proposition that one first must define carefully and analytically the precise purpose of measurement, taking into account the organization setting; then I pointed out that purpose translates to who will make what decision using the measurements. Second, I proposed that purpose and decision should be the key ingredients in determining what will be measured, but I only touched on how one goes about translating purpose into what is measured. I skirted the transition because only gross and tentative rules or guidelines are available. Basically, the measurement specialist must—as a first cut—use his best judgment. Since his best judgment may be wrong or may be severely distorted by cost or other practical considerations, it is essential that the development of a testing program be viewed as a cyclic feedback process, or a cut-and-fit process, with a continual flow of information on whether decisions using test data are good ones. Information on the flaws in such decisions provides the means for changing the measurement and—over time—shaping measurement to maximum support of the decisions that need to be made.

The term "validity" in psychology has many meanings—and the meaning varies depending on the person and on the context in which the term is used. For this reason, I shall avoid these ambiguities and discuss more broadly what one should consider in dealing with the effectiveness of measurement.

The first question to consider is the accuracy of the measurement. What are the tolerances of the emerging numbers?

It is tempting to propose "the more accurate the better." But, that proposal is untenable because cost of measurement increases as requirements for precision increase, in the same way as measurement to one-ten thousandth of an inch is more expensive than measurement to the nearest foot. Just as we decide on tolerances for a length measurement by considering our purpose—whether it is watch-making or road-building—the precision needed in psychological measurement depends on the purpose of measurement, that is, the nature of the decision that is to be made.

The second question regarding effectiveness of measurement concerns stability. If one retested at some later time, how similar would the measurement numbers be to a first set of numbers? Psychologists normally call this characteristic "reliability" but, as with the term "validity," "reliability" has multiple meanings and use of the term is more likely to confuse than to clarify.

How much stability is needed? The hoary tradition of psychological measurement includes the rule that a "correlation of .8 or more is needed for individual decisions; a

correlation as low as .3 can be used for group decisions." This serves as a *general* rule of thumb and, therefore, cannot fit anything. Much better than the all-fitting and hence never-fitting rule is the analysis of purpose and what is to be measured. From analysis of the purpose, one can define the kind of stability of measurement that is needed. From analysis and interpretation of what is being measured, one can distinguish between stability of measurement as it pertains to mechanics of measurement and as it pertains to the nature of what is being measured. In some instances, stability over time would be nonsense, for example. Suppose we administer a typing proficiency test to a group about to begin training in typing. Wouldn't it be foolish to expect test scores secured after training to be about the same as the first set?

The third question under the heading of effectiveness is the pay-off. How much better, in practice, are the individual decisions reached using the measurement than those reached without such information? This question can readily be cast into terms very familiar in industry: How much would it cost to save how many dollars? What is the net gain? However, in order to do such an analysis, it is absolutely necessary—to revert to my main thesis—that the purpose of measurement be analyzed and defined very explicitly, down to exactly what decisions will who make using the measurement data. With decisions defined, it is possible—and, perhaps, even routine—to perform a cost-effectiveness analysis of psychological measurement.

Measurement in industry has enjoyed only mixed success at best, and the question "Is testing worth it?" addressed to management most often results in the answer "I don't know." I think there are two related reasons for this unclear state of affairs.

First, there are many industrial managers who enter internal, deliberative policy councils with a personal conviction that what is really important cannot be measured by tests and that tests and psychologists are not to be taken seriously. In that same council, frequently, will be a testing enthusiast and, after a period of wrangling, the traditional compromise will occur: "Let's try it out on a small scale." Unfortunately, the small-scale approach frequently leads to skipping the crucial steps of analyses to establish purposes to the level of who will make what decisions with the information. Therefore, any hope of getting a good fix on exactly what is to be measured is sacrificed. Usually, a conveniently available test with a name that seems about right and that may have been recommended as a good test is chosen for trial purposes—whether or not it fits the situation and purpose.

Second, exacerbating the instant magic of choosing a convenient test is the fact that, rather than programing a systematic cut-and-fit program for choosing and/or developing measures, a one-shot tryout is undertaken. If the test passes, it's in; if not, testing is out for the company.

Good testing is more expensive than poor testing or no testing. Analysis to determine whether good testing is worth the trouble is not very difficult, once analysis and definition have proceeded to the level of who will make what decision with the information. The costs of poorer decisions in excessive training costs, reduced productivity, or costs of firing someone and hiring a replacement can be estimated, at least roughly. In addition, costs of developing and using a measurement system can also be estimated, at least roughly. From such data, one can calculate a break-even point in terms of the amount of improvement in decisions that is needed to recover costs of measurement. Usually, since training, selection, hiring, firing, and other consequences of decisions are so very expensive, it will be found that even miniscule improvement in the quality of decisions will more than pay for a good measurement program.

Summary

In closing, I should like to repeat my main points:

First, philosophical disputes about whether a person's characteristics can be

measured are pointless. Anything that can be specifically defined can be measured. Such definitions should be in terms of behaviors that can be observed.

Second, to develop a testing program that is useful and cost-effective, the planned use of the test information must be carefully defined: that is, who will make what decisions using the measurements to be obtained.

Third, analysis and interpretation of the particular purpose and the particular industrial setting are essential to decide, hypothesize, estimate, or guess what should be measured. What are usually considered to be different measuring techniques for the same thing are, in fact, measures of different things.

Fourth, the effectiveness of measurement should be evaluated in terms of precision stability, and amount of improvement in organizational activities, all of these considered in terms of the decisions for which measurement provides support. For maximum return on the testing dollar, it is essential to proceed cyclically, continually improving the measurement program in the light of. feedback on how decisions are improved—or not improved—by measurement data.

Fifth, analyses of saving that can be accomplished by improved decisions are usually startling, producing dollar gains far beyond the cost of developing and employing measurement.

My main thesis has been that measurement must be considered in the particular framework in which it is to be used—and here I am talking about measurement in general! I, therefore, call your attention to Dr. Lyons' presentation on quality control, an excellent illustration of the concept of defining who will make what decision using what measurement information.

MEASURING EFFECTIVENESS: QUALITY CONTROL OF TRAINING

J. Daniel Lyons

As the philosopher Seneca observed, "When a man does not know what harbor he is making for, no wind is the right wind." And when training goals have not been precisely defined in terms of measurable on-the-job performance, no training technique is the right training technique. The most pervasive weakness of training programs is lack of precision in locating the harbor of improved job performance. As a result, they are buffeted constantly by the winds of promise and innovation in training—but no wind is the right wind.

Development of new training programs and the introduction of changes in existing programs are fruitless exercises unless and until the means for assessing progress toward precisely defined goals have been developed. Behavioral psychologists have been portrayed by some critics as "drab purveyors of the obvious." In this paper, I may well be adding credence to that observation. It is obvious, is it not, that one does not introduce change unless there exist mechanisms for assessing the effect of the change? I am in the role of a drab purveyor of that obvious and fundamental principle. Because in government, industry, the public schools, and wherever training and educational programs exist, that obvious principle is being continually violated—at a fantastic cost in wasted dollars and human potential.

The process of developing the raw material of human potential deserves a system of quality control at least as carefully developed as that applied to the manufacturing process. By a quality control system I mean essentially an information system and a system of concepts, models, and procedures designed to accomplish four main objectives:

(1) Quality assurance
(2) Control of student progress
(3) Training program improvement
(4) Training system diagnosis and change

The quality assurance function is illustrated in Figure 1.

Does the product meet the specifications? This question cannot legitimately be posed unless and until the specifications have been delineated in terms of operational requirements and these requirements have been reflected in end-of-course proficiency measures. The intent is to rid the training system of criteria based on *amount* of training in favor of demonstrated proficiency in the required job elements. Systematic application of precise job performance criteria through a quality control system results not only in an improved product, but also in the discarding of irrelevant material. Thus, the cost of installing an effective quality control program is amortized through savings in the training program, particularly in personnel time of instructors and students.

The second objective of a quality control system is to provide a means of selecting and organizing the learning experiences of the students to facilitate achievement of the objectives.

The training program depicted in Figure 2 is composed of a series of segments or modules (upper half, Figure 2). Conceptually, these may be as long as a major phase of the course, or as short as a single brief lesson. Each such segment or module is designed to help the student meet specified learning objectives.

Quality Assurance

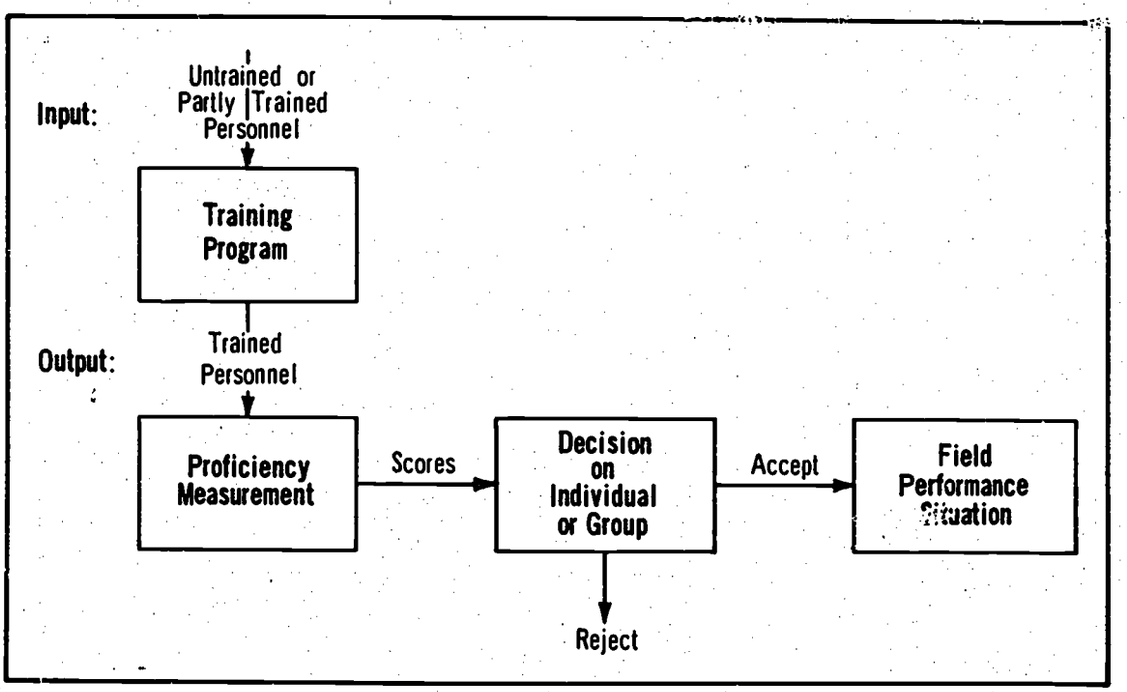

Figure 1

The decision options (lower half, Figure 2), include those of sending the student forward to the next segment of the course, recycling, or giving special corrective training. Generating information to aid in choosing among the options is a function of a quality control system. It should be noted that the option of special corrective training is contingent upon the precision of the diagnostic instrument; that is, the evaluation procedure must be capable of identifying specific weaknesses toward which the corrective training can be directed. The goal is a system by which the trainee is continuously evaluated, selectively corrected, and advanced as performance standards are met, and *only* as they are met.

The first two objectives, quality assurance and control of student progress, are concerned with assessment of student performance. The third objective, shown in Figure 3, is that of program improvement; the emphasis is on program assessment rather than assessment of the individual trainee. Unfortunately, too often changes in training programs tend to be based on administrative edict. We are all familiar with those frustrating situations in which changes in management bring about changes to conform to the biases of the new manager; for example, the shifting emphases on theory and practice in the training of repairmen depending upon the views of upper management rather than job requirements and performance. A systematic quality control process that can identify weaknesses and strengths in the program by assessing and diagnosing the performance of the trainee provides a bulwark against the shifting winds of administrative edict. Further, the control process is necessary in order to assess the effects of changes made to strengthen the program. The most important motivator that can be supplied to any trainer is precise and accurate feedback on the results of his efforts. If this is supplied, training *will* improve, if only by trial and error.

Control of Student Progress

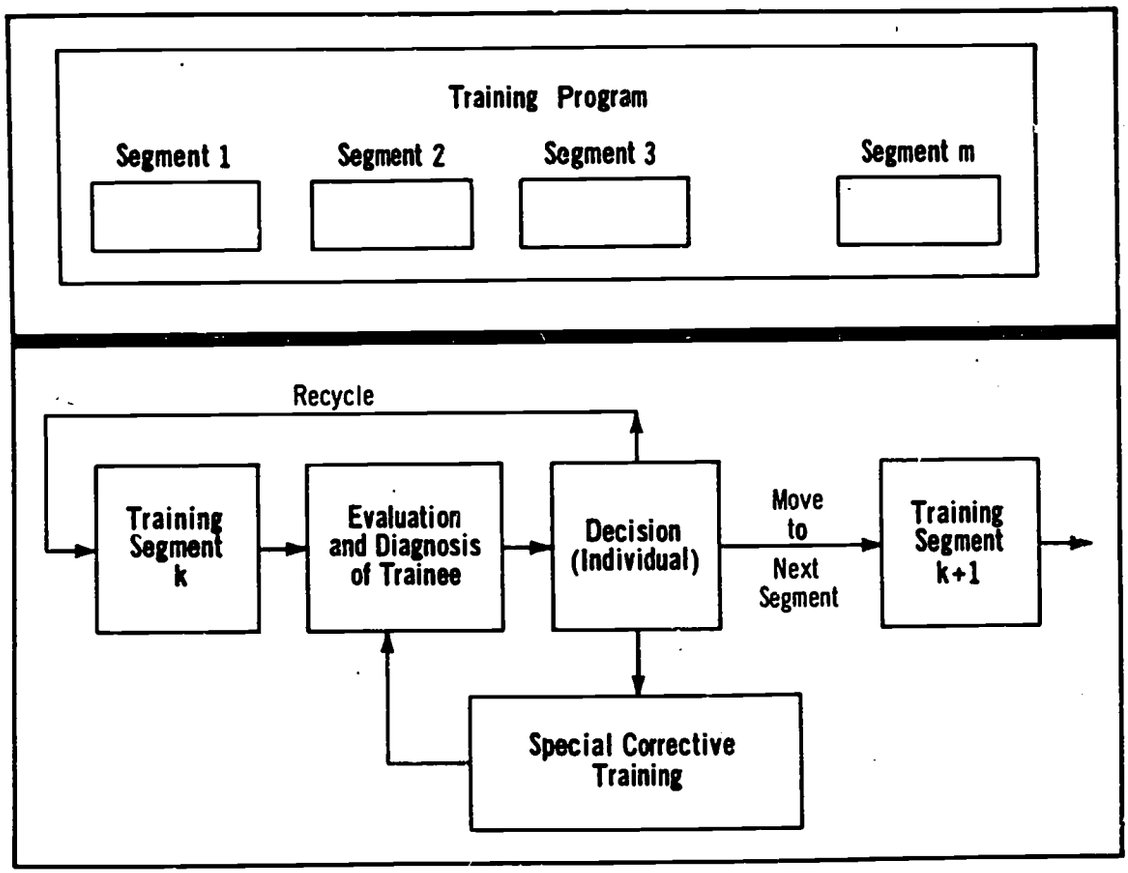

Figure 2

Training Program Improvement

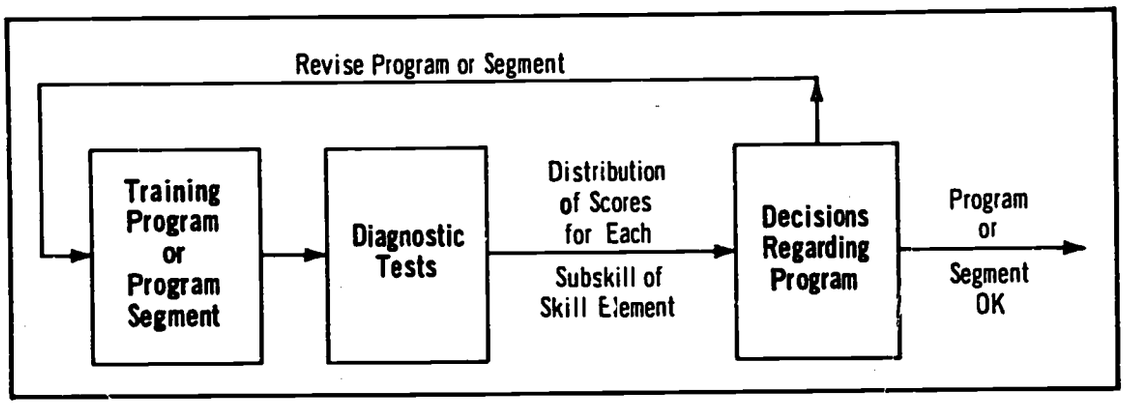

Figure 3

From a Training Director's point of view, Figure 4 may be viewed in the following manner. From the operational elements of the organization, the training system receives performance requirements that are ridiculously inflated or impossibly vague, which must be met with trainees and instructors of minimal aptitude and experience supplied by the Personnel Department, while operating under policies and procedures that are unrealistic, or inflexible, or antiquated, or obscure, or all of these, while utilizing outdated equipment and facilities, and operating on a miniscule budget.

Training System Diagnosis

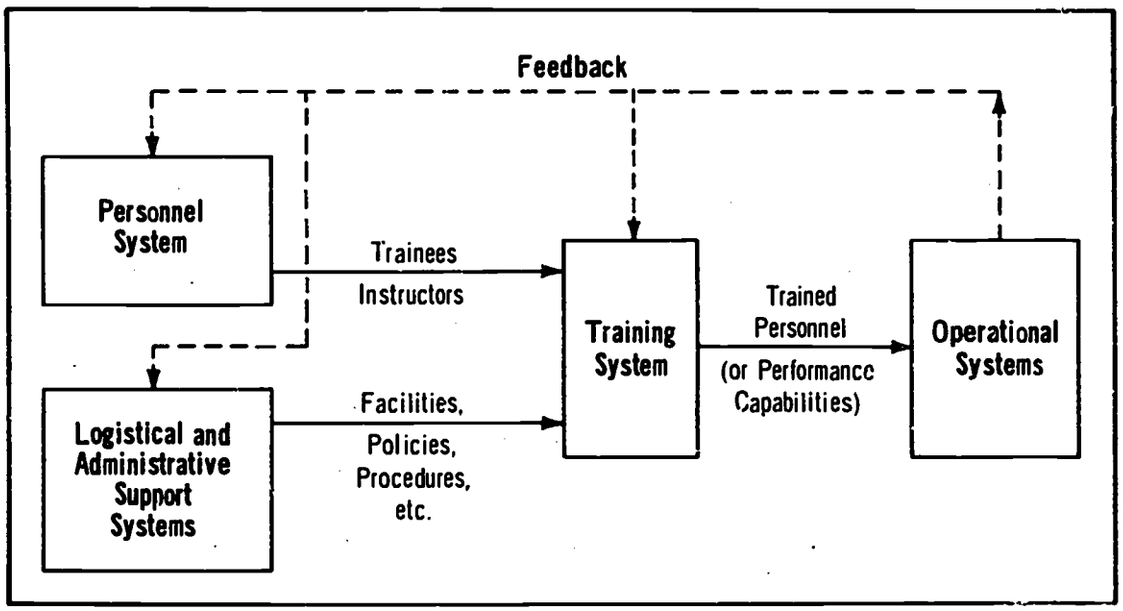

Figure 4

While that may be the world as seen by those of us concerned with training, it is safe to assume that each of the other elements of this system *and* management see somewhat different worlds. An adequate quality control system can alleviate the resulting stresses and strains by providing the information that helps to identify and define the problems and to evaluate the effects of attempted solutions.

The training system is all too often the scapegoat for problems resulting from poorly defined operational requirements, inappropriate utilization of training system products, inadequate personnel selection procedures, and ill-conceived personnel policies. A well-designed quality control system can serve to put our training house in order and provide the basic information for productive interaction with other systems in the organization. In short, it may get the monkey off our backs *or* fasten it there more firmly, if appropriate.

The essential elements of a quality control system are:

(1) Training objectives (performance requirements)
(2) Proficiency and diagnostic measures
(3) Data reduction and analysis
(4) Procedures for decision and corrective action
(5) Communication procedures
(6) Managerial support

For quality control, crucial information derives directly from training objectives. They form the keystone for a useful and effective quality control system by providing not only the specifications for instruction, but also the basis for evaluating instruction. Thus, we must begin with a complete set of good training objectives for a training program, and these objectives represent the mission of the training system.

Management plays the beginning role with regard to training objectives by defining exactly what is to be accomplished by the training system. The raw material for such defining comes from many sources—policies, plans, specifications for new equipment, information concerning on-the-job performance of earlier graduates, information about on-the-job requirements, and so forth.

The management element assembles all such information and decides on *terminal* training objectives. In order for the *terminal* objectives to be most useful, they should be in the form of detailed specifications.

With terminal objectives defined, the training operations element is responsible for developing detailed training objectives and for providing graduates who can perform as defined by management. The set of terminal objectives forms a complete inventory for evaluation. The training objectives also include information about the conditions under which tasks are expected to be performed and thereby define test conditions. Further, the training objectives also include the standards or tolerances for the tasks in terms of accuracy and speed requirements; these are also tolerances for use in scoring an individual's performance on a task.

In order to assess the effectiveness of how the training system is performing, another kind of information is needed about each task—the minimum acceptable percentage of students capable of performing within tolerances. Cost and time aside, it would be desirable for every student to be able to perform every task within the defined tolerances. However, achieving such a goal would be likely to make the cost and time for training intolerably large. Something short of 100% of the students capable of 100% of the tasks must be defined as an acceptable standard of effectiveness of the training system.

The standard must, however, take account of the varying criticality of the tasks. Ninety percent of electricians being 90% correct in the procedures for grounding an electrical circuit during repairs is *not* an acceptable standard. Fifty percent knowing the correct nomenclature of 50% of the contents of their tool kits may be acceptable on a particular job. The criticality measure for any task is basically an assessment of the effect on the operational system of the incorrect performance on that task. In assisting in the development of a training program for stock clerks, we found that the system could absorb, with minor turbulence, an error in the nomenclature of an ordered item but that the stock *number* was highly critical—a misplaced digit could produce an avalanche of toilet paper instead of a fork-lift truck. Similarly, the delivery address was of medium criticality, producing serious delay in delivery—but a misreading of the unit of issue—and we have an avalanche of toilet paper.

The second element, tests and measures, does not make a quality control system—yet they are clearly an essential element of any such system in order to provide the data base on which the system rests. In quality control we are particularly concerned with the diagnostic capability of our testing procedures. We must be able to pinpoint the strengths and weaknesses of the training for each detailed objective as a basis for decision and action to improve or modify the training. In the light of Dr. Cogan's comprehensive discussion of tests and measures, further discussion of this topic seems unnecessary.

It should be re-emphasized, however, that quality control requires absolute rather than relative criteria. Scores and grades must reflect how many of course objectives have been mastered rather than how a student compares with other students. Further, we must ensure that we are not wasting our training time and the potential of our trainees by

failing them for the wrong reason. The key is job-relevance of both training and testing. If the job requirement is to replace the bad part in a TV set on the basis of observation of symptoms, the ability to quote and manipulate Ohm's Law is not job-relevant. Our carefully controlled studies document the fact that many potentially excellent electronics repairmen in a number of training programs have been discarded because of irrelevant weaknesses in physics and mathematics.

The test scores in and of themselves carry little meaning. As a third element, test data must be analyzed and interpreted before they can yield meaningful inputs to decision processes. The data reduction generally involves three kinds of considerations—central tendency, variability, and stability. The central tendency is calculated to show the overall performance of the group—average, mean, or perhaps, more useful, the percentage of a class able to perform each specific task at or above the minimum standards. The variability or spread is generally characterized by calculating the standard deviation, while stability is identified by the standard error in order to distinguish the accidental or incidental deviations from those that have a "real" basis.

In the analysis of the data that have been reduced to measures of central tendency, variability, and stability, three basic questions arise regarding performance on each task. First, how does the central tendency compare with the standard? Has the class performed above, below, or at the standard? Second, does the class performance fall within tolerances established for the standard? Third, how critical or important is the task to operational performance? As indicated earlier, the criticality of the task has direct implications for the urgency of corrective action. The criticality dimension is built into the analysis by differential standards and tolerances for specific tasks.

The collection, reduction, and analysis of the test and performance data are necessarily designed to support a program of corrective actions, the fourth essential element of the quality control system. It is, unfortunately, almost commonplace to find massive collections of training data, created at considerable effort and expense, lying idle. Too often such data are assembled without a specific plan for utilization or in the absence of specific procedures for implementing the existing plan. Prior to the collection and analysis of the data, there must be procedures for corrective action—that is, specification of the process by which decisions are made and means of assigning responsibility for implementing the actions selected. These procedures should be designed to identify problems and to assign priority to their solution. The highest priority for action is for those cases where the data analysis shows that performance is seriously out of the tolerance range.

In order to maintain confidence and support of management and of the operating elements, it is important that such problems be identified by the training element and corrective action initiated immediately. The system should act rather than react to external complaints. A complete action program should include procedures for:

(1) Identifying points and places where something seems to be seriously out of tolerance and immediate action is indicated.

(2) Identifying points and places that are "suspicious," and that warrant investigation as time and resources permit.

(3) Establishing a normal routine work load for continuing study of the training program when everything is going well.

Obviously a quality control system must include carefully designed communication procedures. The information generated by the system must be differently packaged for transmission to the responsible individuals on an appropriate schedule so that the necessary decisions can be made on a timely basis. Equally important are provisions for flow of relevant information into the system—changes in operating procedures, new equipment, modifications in personnel selection procedures, policy decisions affecting training, and so on.

Proper communication is vital to maintaining managerial support, which is both a cause and an effect of a dynamic quality control system. The quality control system cannot operate effectively without strong support from all managerial levels, nor will this support continue unless the system operates effectively. Support from management is especially needed, because the data produced by the quality control element may be unpleasant. However, if the information is directed toward corrective action, quality control can be viewed as the shared mission of management and the training element: producing the tangible asset of a well-trained addition to the company work force.