Catherine M. Sleezer
Editor

Improving Human Resource Development Through Measurement

Third in the Theory-to-Practice Monograph Series

Contributors:

J. Diane Beezer-McClure
Dale C. Brandenburg
C. Victor Bunderson
Gebreyesus Hamda
C. W. Hoover Jr.
Randolph I. James
Jospeh T. Martelli
Sandra J. Merwin
Nancy R. Mosier
Brian P. Murphy
Jeffrey D. Oberlin
Gene W. Poor
Suzanne Saxe
Scott A. Sayre
Richard A. Swanson
Michael T. Wood

Published by the
University of Minnesota
Training and Development Research Center
Department of Vocational and Technical Educat
and the
American Society for Training and Development
Research Committee

Human Resource Development Theory-to-Practice Monograph Series

This monograph is a joint project of the Training and Development Research Center of the University of Minnesota and the Research Committee of the American Society for Training and Development. It is the third in a series of theory-to-practice symposiums and monographs that involve both distinguished scholars and human resource practitioners. The symposium that led to this monograph was held on the University of Minnesota campus in St. Paul, February 28–29, 1988.

The first two symposiums held at the University of Minnesota resulted in the following monographs:

Swanson, R.A., & Gradous, D., Eds. (1987). *Human resources and organizational change.* Alexandria, VA: American Society for Training and Development.

McLean, G.N., & DeVogel, S.H., Eds. (1988). *The role of organizational development in quality management and productivity improvement.* Alexandria, VA: American Society for Training and Development.

Published by: AMERICAN SOCIETY FOR TRAINING AND DEVELOPMENT

1630 Duke Street
Box 1443
Alexandria, VA 22313

703-683-8100

Contents

Measurement Research

Measurement Practice Meets Measurement Science

Richard A. Swanson
Professor and Director
Training & Development Research Center
University of Minnesota

Catherine M. Sleezer
Training and Development Consultant
Minnetonka, Minnesota

Almost everything that is important to American business and industry leaders is measured. Human resource development (HRD) practices, however, are seldom systematically measured and evaluated. One reason is that HRD has a long tradition of not being taken seriously by upper management and not being expected to provide effectiveness data (Swanson & Sleezer, 1987). Upper management has not been concerned about measurement in HRD because the resources allocated to HRD have been relatively small compared with capital investments and because HRD has been viewed as a cost rather than an investment. Consequently, non-HRD managers have believed that the added expense of closely monitoring the relatively small HRD investment would not be cost-effective, and, in the place of systematic measurement, they have based HRD decisions on secondhand information.

But the perceptions of these same managers are now changing. More and more they are recognizing that the way a company develops and works with its personnel is one of the keys to organizational and financial success. Consequently, pressure to evaluate HRD program effectiveness is now coming from non-HRD managers at the highest organizational levels. Upper management is increasing its investment in HRD and is asking for evidence of effectiveness. This view of HRD as an investment with the potential of significant financial returns (Spencer, 1986; Swanson & Gradous, 1988) is challenging the traditional view of HRD as expensive overhead that must be cost-controlled.

Most HRD scholars and practitioners would agree that measurement and evaluation in the HRD profession lag far behind the available knowledge of best practices (Parker, 1986). The irony of the situation is that the lag between measurement theory and practice has been perpetuated by HRD professionals who have accommodated this view in their organizations and who have not challenged upper management's traditional views and who, for the most part, have restricted their measurement activity to soliciting satisfaction ratings from participants.

The SLP Evaluation Model

The traditional Satisfaction, Learning, and Performance (SLP) Evaluation Model (see Figure 1) for HRD/training is theoretically sound, yet rarely seen in practice (Swanson & Sleezer, 1987). It focuses on measuring an HRD program to make summative decisions about the overall effectiveness of the program, the program facilitators, and the individual participants.

Figure 1—SLP Evaluation Model: Categories and measures

Satisfaction	Learning	Performance
• Participant's Satifaction • Participant's Supervisor Satisfaction	• Knowledge Test • Skill Test	• Job Performance • Financial Performance

The *satisfaction* component of the SLP Evaluation Model focuses on how pleased the participants and their supervisors are with the program. Participant satisfaction is almost universally used to measure HRD programs, although supervisor satisfaction is usually not measured on a systematic basis.

The *learning* component of the SLP Evaluation Model is connected to the quality of HRD/training experiences. The traditional HRD mandate is that participants "know it" and "be able to do it." Thus, measures that verify these capabilities and simulate actual work are important measures.

The *performance* component of the SLP Evaluation Model concerns the context and workplace in which the participants act upon their new expertise. Behavior in the workplace (or resulting units of performance) and financial benefit analysis are the two bottom-line indices of added value to the organization (Swanson & Gradous, 1986, 1988).

Plenty of evidence shows that training participants can be highly satisfied with programs, yet have learned very little and not perform any differently back on the job (Dixon, 1988). There is also plenty of evidence that trainees can learn a great deal during training, yet not necessarily perform any better back on the job. In these instances, training usually has been applied to a nontraining problem. This monograph addresses the importance of accurately identifying an organization's critical performance needs that require training or other HRD interventions and the importance of implementing high-quality programs that meet those performance needs.

Theory and Practice

Dr. C. Victor Bunderson, Vice President for Research Management, Educational Testing Service, is the distinguished scholar for this monograph. He talks to us about measurement science within the *learning* component of the SLP Evaluation Model. He focuses on how learning efficiency and effectiveness can be increased by applying new measurement science knowledge. He challenges practitioners to alter radically their view of measurement to the point of affecting the design and delivery of training. He goes beyond summative evaluation and into formative evaluation, and describes a method for increasing the efficiency and effectiveness of the training-in-process through the use of measurement tools.

Responses to Bunderson's chapter from the distinguished practitioners are varied. The responses are grouped into three follow-up sections:

- Section 2: HRD and the Training Function
- Section 3: Measurement in Business and Industry
- Section 4: Implementation of HRD Measurements

In Section 2, Randolph I. James, James & Associates, frames training within the larger context of organizational performance. J. Diane Beezer-McClure, Honeywell, Inc., then discusses the importance of front-end analysis for training programs that are designed to solve organizational problems. Completing this section, Gabreysus Hamda, Florida Department of Transportation, zeros in on the role of training within HRD.

In Section 3, Jeffrey D. Oberlin, Motorola, looks at measurement in the business system and distinguishes between organizational and individual performance. Next, Charles W. Hoover, AT&T, captures the changes in the workforce that demand new knowledge, training, and measurement. Michael T. Wood and Joseph T. Martelli, Industrial Technology Institute, teamed with Nancy R. Mosier, Arthur Andersen & Co., to deal with the difficult issue of integrating the measurement and training systems into the organization.

In Section 4, Scott A. Sayre, Sayre Media Consultants, describes technological alternatives that can be used in applying Bunderson's theory. Gene W. Poor, Good Display, Inc., and Brian P. Murphy, The HRD Department, Inc., take on the practical issue of test anxiety. Next, Suzanne Saxe, Wells Fargo Bank & Co., focuses on the practical issue of measurement given limited resources. Dale C. Brandenburg, University of Illinois, outlines critical issues for professional practice. And Sandra J. Merwin, Sandra Merwin Enterprises, concludes Section 4, and the monograph, with a set of ethical concerns and guidelines for action.

References

Dixon, N. (May 1988). *The relationship between responses on participant reaction forms and actual learning in a training setting.* Paper presented to the HRD Professor's Network, Dallas.

Parker, B.L. (1986). Summative evaluation in training and development. *Journal of Industrial Teacher Education, 23*(2), 29–55.

Spencer, L.M., Jr. (1986). *Calculating human resource costs and benefits.* New York, Wiley.

Swanson, R.A., & Gradous, D.B. (1986). *Performance at work: A systematic program for analyzing work behavior.* New York: Wiley.

Swanson, R.A., & Gradous, D.B. (1988) *Forecasting financial benefits of human resource development.* San Francisco: Jossey-Bass.

Swanson, R.A., & Sleezer, C.M. (1987). Training effectiveness evaluation. *Journal of European Industrial Training, 11*(4), 7–16.

Measurement Science and Training

C. Victor Bunderson
Vice President for Research Management
Educational Testing Service

Can a revitalized measurement science help solve the nation's training problems? The need for training and retraining is currently a central element in discussions about the economy of the United States. Our country's ability to compete in world markets is being challenged severely by the willingness of workers in other countries to work for lower wages, and by their increasing ability to mass produce quality goods. As a result, the U.S. economy's ability to maintain a high standard of living for all Americans is increasingly doubtful. Many companies with jobs requiring minimal education and little ability to respond flexibly have automated or sought offshore labor. The jobs that remain require higher levels of technical knowledge and skills, yet the pool of educated and skilled people is growing smaller. A large, poorly educated underclass, which includes many minorities, is increasing. Not only is our standard of living undermined through decreased economic capacity, but the democratic foundations of American society are threatened because democracy requires an educated citizenry.

We must educate the young and prepare them for increasingly complex and technical roles in a world economy. We must retrain adults, on the average, eight times during their careers. High levels of competence are needed. As pointed out in the recent report by the Carnegie Forum on Education and the Economy (1986), the alternative to the old model of using semiskilled labor is to

> revise our view of the role of the worker in the economy. In the future, high-wage level societies will be those whose economies are based on the use on a wide scale of very highly skilled workers, backed up by the most advanced technologies available (p. 13).

Alan Greenspan (1988), chairman of the Federal Reserve Board of Governors, spoke last year about America's ability to function in a global economy. He stressed the intellectual component as key to the rise of the GNP since the turn of the century and certainly key to any future rise: "In addition to modernizing the physical capital stock, we must also concentrate on broadening and deepening our 'human capital' " (p. 8). He also stated,

> I am referring not only to those [skills] that are specific to particular jobs and thus can quickly become obsolete. Rather, I want to stress the need to acquire broad analytical and problem-solving capabilities that will facilitate the processing of information and enhance one's ability to adapt to the demands of a complex, dynamic economy (p. 9).

The issue addressed in this chapter is whether a revitalized measurement science can be of significant help in achieving new levels of effectiveness in the kinds of training the United States will need to be competitive in the world. The science of educational measurement deals with that elusive, valuable, and invisible quality—human competence. Measurement scientists have struggled with the problem of measuring higher levels of competence, including analytical and problem-solving skills. One product of this struggle has been tests given under standardized conditions with the purpose of making visible that hard-to-observe but highly valuable resource, high-level competence. However, measurement science as it has functioned in the past is not likely to make a significant impact on the problem of achieving high levels of technical competence through retraining the American workforce. One reason is that it has not dealt with measuring learning and improvement but rather with predicting future events from current scores. Another is that the conventional models of expertise have been too simplistic. Measurement science itself is now in the throes of change in response to new interactive technologies and the impact of new scientific disciplines. These emerging changes have the potential to revitalize measurement science and make it more relevant to training practice.

The new technological alternatives that affect the development of human competence include interactive training delivery systems. These delivery systems employ computers interfaced to mass memories containing large files of video images, audio messages, computer-generated graphics, and computer-generated sound and text. The systems provide not only a wide variety of formats for information display, but also new response modes as alternatives to the standard multiple-choice format. These include keyed responses employing natural language and mathematical symbols, touch screens, and the use of cursor keys or a mouse to point to screen locations and "move" objects from one place to another on the screen. Interactive training delivery systems, however, have not reached their full potential for many reasons. The reasons include cost and lack of wide distribution, but most fundamentally, inadequate understanding of learning processes, teaching processes, and how to overcome organizational inertia.

Better scientific foundations can help reduce the problems that have slowed the use of interactive training systems. Increased scientific understanding is currently emerging from new disciplines such as the cognitive sciences, instructional science, and applied artificial intelligence and from studies of organizational change. It is through disciplines such as these, and the use of new delivery systems, that measurement science is being revitalized. At Educational Testing Service, we are eagerly engaged in this revitalization process. The Research Division at ETS has scientists involved in all the areas mentioned above and in the use of new interactive delivery systems.

This chapter emphasizes the interplay between training and a measurement science revitalized by new scientific methods and technologies. In addition to discussing how measurement science can improve training, it examines how the problems and perspectives provided by interactive training systems are stretching measurement science in new ways and producing a cross-fertilization that should expand both disciplines greatly. The following discussions embrace (1) measurement science concepts and definitions and (2) a description of a future training scenario in which these concepts are applied.

Concepts of Measurement Science

Two measurement science definitions useful for training professionals are educational measurement and performance measurement. *Educational measurement* is the process of specifying, for educational purposes, the positions of persons, situations, or events on educationally relevant scales under stipulated conditions. A related concept is *performance measurement*, which is a similar process of specifying the positions of persons, situations, or events on scales relevant to valuable (generally, economically valuable) accomplishments under stipulated conditions.

Solutions to training problems will have to be delivered by an applied science of training. The issue addressed in this chapter is: To what extent can this applied science be improved by better measurement? Measurement scientists must rise to a new set of challenges in integrating measurement into training and improved performance in a way that has not been accomplished to date.

Common Applications of Measurement in Training and Performance

Admissions, selection, certification, and licensing

To date, the more sophisticated contributions of measurement science have not been an intimate part of training or education. Instead, the most common applications of measurement science include the use of carefully crafted admissions tests for selection of applicants into colleges or work assignments and the use of testing for certification or licensing in an occupation. Powerful approaches for measuring the growth of human competence as it develops over time as a result of learning have not yet been a contribution of measurement science. The approach that comes closest is a combination of placement tests and diagnostic tests given before study begins, and proficiency tests administered midway and at the end of the course.

Counseling and guidance

Tests of ability, interest, personality, and other attributes are currently a substantial part of the process of counseling and guidance, whether for career planning, college admissions, employment counseling, or therapeutic purposes. These will continue to serve important needs but do not affect training directly.

Performance engineering

The "educational measurement community" consists of professionals in the traditions of Thorndike, Thurstone, and others who have focused on measuring aptitudes and knowledge, and on the psychometric and statistical issues of scaling, equating of scales, validity, reliability, and so forth. They constitute a group that is very different from the professionals who have focused on performance engineering. This group developed out of the fields of human factors (earlier known as "human engineering") and job training. Although the initial focus of human engineering was designing equipment to be compatible with the capabilities and limits of people, the field has broadened to include designing jobs and specifying the needed human skills. Performance engineering now affects both selection and training, as well as the design of hardware and software systems.

Measurement is an integral part of performance engineering. A 656-page textbook commissioned by Bell Laboratories in 1978, *Human Performance Engineering: A Guide for System Designers* (Bailey, 1982), is replete with examples of the use of measurement in assessing human performance both globally, as it relates to specific jobs and tasks, and, specifically, as it relates to human limits and differences in sensing, perceiving, responding, processing, problem solving, decision making, and so forth. In his book *Human Competence,* Gilbert (1978) shows how to use performance measurement as a fundamental instrument for producing high-level competence. His methods can be used for developing models of performance with economic consequences, for evaluating the payoff of competence, and for determining the cost of lack of competence. As yet, the applied fields of performance engineering have not found the advanced statistical and psychometric models of educational measurement to be necessary in their work. Nor have they given much emphasis to the validity of inferences from scores. This chapter argues that by seeking the common goal of achieving higher levels of competence, the sister fields of educational measurement and performance engineering may each benefit greatly from the findings and methods of the other.

Formative evaluation

Applied measurement has been used in formative evaluation for a variety of training-related purposes. In formative evaluation, data are used to improve a system as it is being developed or implemented. Surveys often are used to obtain information during the development of training programs. Survey data can help in refining the job analysis and in determining the most critical tasks that should constitute a training program. Item analysis can be used to improve evolving training programs through formative testing of exercise units. Measurement methods can make a much larger contribution to formative evaluation of training than they have in the past.

Summative evaluation

Summative evaluation uses data to make an overall judgment about the value of a program. Policy makers frequently seek summative evaluations of training programs either as a way to prove the programs are achieving good results or as a way to attack current training policies. When policy makers have positions at stake, the validity of inferences made from summative measurements is frequently strained. It behooves measurement science to offer scales and guidelines that can be used to support valid inferences and interpretations. Also, measurement experts need to be persistently persuasive in encouraging users to meet the conditions for the proper use and interpretation of measures.

In summary, educational measurement has made contributions at the beginnings and endings of learning, but has not contributed much toward improving the process of learning. Applied fields like performance engineering and interactive training are making important contributions to the processes of human learning, but they should be challenged to consider measurement issues more seriously.

Validity: The Foundation of Measurement Science

When measurement scientists ask whether what is supposed to be measured is in fact being measured, they are addressing the issue of validity. The validity of inferences from measures of competence is a key issue in training. This is particularly true of inferences about the status of trainee growth and what can be done to smooth the path toward their achieving higher levels of competence. Interactive technologies (such as computer-controlled videodiscs) offer promise for providing job-like simulations that can be used to measure performance during training. Unfortunately, most uses of these technologies have depended primarily on face validity—the inference that the resemblance of a simulation to an actual job means that performance in the simulation is equivalent to performance on the job. Technology-based performance measures will be strengthened by the evidence of criterion-related validity, whereby the performance levels demonstrated in the simulation are validated against on-the-job criteria. Criterion-related validity is present in flight simulators, where simulator training is followed by actual flying hours. Unfortunately, too many training simulations are not validated against real job-performance criteria.

As important as criterion-related validity is, construct validation may be the greatest contribution that measurement science can offer to training and performance. The term *construct validity* refers to the accuracy of inferences that are based on an understanding of the fundamental constructs (ideas derived from a model or theory) of the domain or task. Scales generated from test items (through factor analysis, for example) can be useful in understanding the dimensionality of the domain. If four scales are needed to deal with the tasks in the domain, the constructs that explain what each dimension represents may lead to inferences about how to train more effectively. The constructs that explain increasing competence, as an individual progresses from novice to expert, are potentially the most powerful contributions of measurement science. Although these constructs can be developed from cognitive or instructional science, they must be validated by measurement science.

The course toward true mastery is seldom smooth and continuous. There are discontinuous changes in the cognitive processes and the mental models used by novices, advanced beginners, adepts, and experts. If measurement can highlight and reveal the nature of these mental models, this information then can be used to refine the specific presentations and feedback in a program of instruction and can potentially speed learning.

Scales, Construct Dimensions, and Anchors: An Example

A major contribution of measurement science has been the development of psychometric and statistical models. These, in turn, have led to scales that span wide ranges of human proficiency. One example is the adult literacy scales of the National Assessment of Educational Progress (NAEP) (Kirsch & Jungeblut, 1986). In this study, the domain of adult literacy was sampled, using simulation tasks that resembled what adults do in the real world. Trained observers administered the adult-literacy items one-on-one to a representative sample of 3,600 young adults, ages 17 to 24. The subsequent analysis produced three unidimensional scales (Kirsch, 1987): prose literacy, document literacy, and quantitative literacy.

- Prose literacy: knowledge and skills needed to understand and use information from texts, including editorials, news stories, poems, and the like;
- Document literacy: knowledge and skills needed to locate and use information contained in job applications, payroll forms, bus schedules, maps, tables, indexes, and so forth;
- Quantitative literacy: knowledge and skills needed to apply, either sequentially or alone, arithmetic operations embedded on printed materials, as in balancing a checkbook, figuring out a tip, completing an order form, or determining the amount of interest on a loan.

Tasks representative of these three types of literacy were scaled using item response theory (IRT). IRT is a mathematical model for estimating the probability that a person will respond correctly to a particular task. To determine this probability, analyses within a given scale were carried out in two steps. First, the parameters of the tasks were estimated, which, for the NAEP, included item discrimination, item difficulty, and, where appropriate, guessing. Second, levels of proficiency were estimated for individuals and groups. The tasks, which anchor points along the scales, provide a criterion-referenced interpretation of the corresponding points. The group proficiency estimates provide data for norm-referenced interpretations (Kirsch, 1987).

Supplementary information to help interpret performance at various levels along each scale was developed by selecting benchmark tasks (also called anchors) and identifying underlying characteristics that contribute to task difficulty. For example, tasks on the document scale seemed to vary as a function of three characteristics: (a) the number of features or categories of information that the reader had to locate or use, (b) the relationship of the wording in the question to the wording in the document, and (c) the number of distractors or potentially correct answers in the material.

These analyses suggest that in many instances, literacy tasks require the reader to apply complex information-processing skills and strategies to a variety of linguistic structures. Becoming aware of these complexities deepens our understanding about the nature of literacy. Furthermore, it allows us to characterize the nation's literacy problem in terms of levels of performance in employing these skills and strategies, rather than in terms of overly simplistic labels, such as "illiterate" or "functionally illiterate."

Moreover, gaining a better understanding of the underlying skills and strategies associated with performance at various levels on each of the scales has important implications, not only for building more reliable and valid literacy measures, but also for designing instructional programs. Adult literacy programs too often are based on models suitable for elementary schools, where reading, for the most part, is restricted to the use of narrative texts. Results from the NAEP study suggest that emphasis on a single aspect of literacy may not lead to the acquisition of skills and strategies needed to participate fully in society.

The quantitative models of measurement science, in the form of IRT and its variations, made the NAEP analysis possible. The scientific fruits of this large-scale measurement activity are continuing. Educational Testing Service, which administers and analyzes the national assessment, is making a substantial investment toward finding ways to use the NAEP scales and constructs to improve adult literacy. The goal is to develop systems of integrated assessment and instruction embodying *Continuous Measurement*, a new generation of computerized measurement explained in the next section.

Concepts of Measurement Science Expanded by New Technologies

Recent developments in computerized measurement can be summarized by placing them in a four-generation framework (Bunderson, Inouye & Olsen, 1989). Each generation represents an application of measurement science through technologies of increasing sophistication and power. The generations are as follows:

- Generation 1, Computerized Testing (CT): administering conventional tests by computer.
- Generation 2, Computerized Adaptive Testing (CAT): adapting the test to the individual test-taker by selecting each succeeding task on the basis of the test-taker's performance on previous tasks.
- Generation 3, Continuous Measurement (CM): using calibrated measures embedded in a curriculum to estimate changes in each learner's proficiency continuously and unobtrusively.
- Generation 4, Intelligent Measurement (IM): introducing knowledge-based (artificially intelligent) computing to the decision-making processes of computerized measurement. Knowledge-based computing becomes part of the subprocesses of educational or performance measurement.

The first generation: Computerized testing (CT)

The CT generation is typified by translations of existing tests to computerized formats or by new, non-adaptive tests that are similar to manually administered tests but utilize computer capabilities for all or most test administration processes.

The first generation offers many advantages to the measurement processes, not the least of which are immediate scoring, immediate feedback, and testing on demand. CT generally proceeds faster than testing using pencil-and-paper formats and makes possible new response formats. These formats include typed responses in the form of numbers, equations, words, or phrases; pointing responses through the use of touch screen, cursor keys, or mouse; and auditory responses. Joy sticks, track balls, steering wheels, and other devices offer additional modes of response. Also, many new displays are available, including motion pictures, animation, randomly accessed audio, and computer-generated overlay on video images. With the computer, it is easy to measure the latencies of individual responses and elapsed times for complex sequences of actions. It is also possible to vary dynamically the length of time that displays are presented. Data collected from computerized tests can be transmitted electronically to a central site for statistical analysis, calibration, scale equating, and various analyses for the purpose of test improvement.

CT enables first-generation tests to be more job-like, hence less decontextualized. Such tests may be less formidable to the test taker, not only because they can be constructed to resemble familiar things in the work situation, but also because computerized test items appear one at a time, not as a formidable booklet filled with page after page of test questions.

The second generation: Computerized adaptive testing (CAT)

The CAT generation of computerized educational measurement is typified by computer-administered tests in which the presentation of the next task, or the decision to stop presenting new tasks, is adaptive. That is, calculations based on the test-taker's performance on previous tasks determine the presentation of each new task. A task may be an item or a complex standardized situation involving one or more responses.

Several types of information may be involved in adapting CAT tests. The most common adaptation is based on an estimated scale value. That is, on the basis of previous responses, an estimation of the position of the learner is made on the underlying latent trait scale. (A latent trait is both the statistically defined scale and the inferred construct explaining score differences.) This estimate guides the selection of an easier or more difficult next item. CAT tests generally use item response theory, which permits both test items and test takers to be identified by their positions on the same underlying latent trait scale. As the estimate of the test taker's position becomes more certain, items closer to the estimated scale value can be presented to the learner, thus making the estimate of the learner's position on the scale more and more precise. Accurately estimating the test taker's position on the scale may be accomplished with 30 to 60 percent fewer items than with a paper-and-pencil test, leading to further reductions in testing time over CT.

Computerized adaptive tests are excellent for identifying a learner's exact scale position on an underlying latent-trait scale. They are not ideal for determining whether a learner is above or below the cut point used for determining certification, licensing, or admission into some program, because in such testing applications it is not necessary to determine the exact scale position of each test taker. Also, the content specifications of the test may be violated if only a small subset of items is administered. Another form of adaptive test, called the Computerized Mastery Test, can be used to optimize cut-score decisions. It adaptively selects well-specified item clusters to sample the content specification adequately and to decide quickly whether a test-taker's score is above or below the cut score.

Computerized tests may adapt with regard to response time by varying the speed of presentation or by requiring faster and faster responses. Such tests may converge on the individual test taker's characteristic speed of perception or speed of response under stipulated conditions.

Another form of adaptation is to adapt the content and the sequence of displays, as in a simulation. Flight simulations, for example, are adaptive in that the parameters of the simulation change in response to the trainee's engagement of the simulation controls. As scientists define the constructs that are operating in given simulations and relate them to the underlying processes, construct-valid scoring schemes will be developed. These scoring schemes will produce multiple scores from the same simulation with which to inform learners and instructors about the learner's performance and how it may differ from the performance of more expert simulation operators.

The third generation: Continuous measurement (CM)

CM uses calibrated measures embedded in a curriculum to estimate dynamic changes in each learner's proficiency continuously and unobtrusively. The tasks used as the units of measurement may be items, item clusters, exercises, unit tests, or independent work assignments. Changes may be observed in the amount learned, in proficiency on different tasks, in the trajectory through the domain, and in the student's profile as a learner.

CM is the first generation of educational measurement that holds the promise of being fully responsive to the needs of learning and instruction. It is integrated into an instructional system and is designed to provide continuing feedback for use by both learners and instructors. Its purpose is to guide and facilitate growth on several underlying mastery scales. These mastery scales have the same properties described earlier for the adult literacy scales. They always span all, or a coherent part, of a domain of expertise. For example, adult literacy was spanned by the domain of simulation items selected for the NAEP Young Adult Literacy Assessment. The domain of interest may be collapsed into a few key constructs, as with the prose literacy, document literacy, and quantitative literacy constructs. These constructs are anchored by demonstrable performance on reference tasks along the way to mastery, so that context and specificity can be given to positions on the scale, rather than simply confronting the learner with a numerical score. Indeed, the numbers do not have to be communicated to the learners and instructors—just the scales, the different types of tasks on each scale, and the complexities associated with mastering the constructs that lead to success on the tasks.

Continuous measurement is unobtrusive. It is not segregated from learning as are midterms, finals, or periodic tests in training courses. Instead of imposing obtrusive and formidable forms of testing, CM resembles that friendly form of testing we have all experienced when our teachers and fellow students discuss correct and incorrect answers on a very recent quiz so that all can learn from the experience. It is diagnostic and advice-oriented to help each learner achieve a desired goal.

Continuous measurement is a key component of an integrated system of instruction and assessment. *Mastery Assessment System* is one name given to the assessment component of an integrated instructional system that allows one to estimate the learners' positions, track their trajectories on the underlying scales, and make inferences about the deeper constructs fundamental to those scales. It is discussed in detail later.

The fourth generation: Intelligent measurement (IM)

IM is defined as the application of knowledge-based computing to any of the subprocesses of educational or performance measurement. IM introduces knowledge-based (artificially intelligent) computing into the decision-making processes of computerized measurement. The potential applications of this capability are virtually unlimited. Three applications of immediate interest are (a) intelligent scoring of constructed responses, (b) intelligent interpretations of score profiles, and (c) intelligent prescriptions for appropriate instruction in the continuous measurement environment.

Computerized scoring of constructed responses would be a major boon to measurement. Currently, labor-intensive methods of holistic scoring are employed in grading writing exercises, pieces of art work, or architectural drawings. Human judges grade the productions after having been trained on a set of standards illustrated and anchored by sample productions representing various points on the rating scale. For example, six essays might be provided, anchoring the attributes of ratings from one to six. Researchers have examined the possibility of computerized grading of computer programs written in the Pascal language (Bennett et al., 1988). The Advanced Placement program for computer science requires such programs to be written, and many are submitted on floppy discs. Bennett and his colleagues are working with Elliot Soloway of Yale University to examine the applicability of his artificially intelligent program called PROUST (Johnson & Soloway, 1987) to the scoring of these programs. Although there is promise in these methods, early applications are not imminent.

Intelligent interpretation of score profiles is a second IM example. In this application, the expertise of human counselors, who advise individuals based on their score profiles, is captured by knowledge engineers and placed in a knowledge base so that the computer may go through the same reasoning as an experienced counselor and generate similar interpretations. Such expertise in intelligently interpreting the profiles could then be replicated at many locations.

The most complex application of intelligent measurement would be to generate by computer the prescriptions, hints, or prompts to advise learners who are moving through a learning system employing continuous measurement. Scores from multiple tasks along the way could be obtained to provide a profile of the characteristics of the individual learner as well as the learner's trajectory on the underlying scales. Human experience with such advanced systems needs to be developed before this expertise can be captured and used in intelligent continuous measurement.

Concepts of Mastery Assessment Systems[1]

The concept of mastery assessment systems, a CM generation concept, was developed during 1986 and 1987. Two multi-year research projects were initiated to develop mastery assessment systems, one in adult literacy and the other in middle-school science (Lipson, 1987).

Two features of mastery assessment systems may be noted at the outset. First, a mastery system is a part of a larger learning/performance information system for a specified curriculum but is not itself a curriculum. Second, the term "mastery" does not refer to minimum competence alone.

Footnote

[1]This material is adapted and updated from discussions of mastery systems first described by Forehand and Bunderson (1987a, 1987b), and similar material found in Bunderson, Inouye, Olsen (1989).

A mastery system as part of a learning/performance information system

Learning/performance information system is the term used to describe the management system that coordinates curriculum materials, instructional resources and schedules, and mastery assessment components into a single, integrated system. The mastery system is the measurement part of the learning/performance information system. It provides assessment information about how learners or trainees are progressing from novice to expert and how their performances compare with the exemplary performance we would attribute to "masters" of the subject being assessed. The total volume of instructional materials, management data, and applications software that occupy the databanks of the learning/performance information system may be much larger than the mastery assessment system materials.

The learning/performance information system is housed in networked work stations communicating to a central file server for data management. This system may be operated primarily for training purposes in a laboratory or learning assistance center, or the system may be more work-oriented. It may be used by a work group in a local area network to automate many of its activities. In the work-group application, the mastery assessment materials and the curriculum materials comprise a lesser part of the system, which is devoted primarily to supporting the work to be automated in the work group.

Two functions of a mastery system within curriculum

A mastery assessment system is not itself a curriculum, but it provides two important functions when integrated with a curriculum. The word "curriculum" is ill-defined. It may be used at different times to refer to the historical evolution of the instructional content and courses, all of the courses in a catalogue, or the materials of instruction used in a single course of study. The Latin root of the word "curriculum" means to run, as to run a course.

A training course is designed to run the learners along an instructor-selected path through some domain of knowledge and expertise. The domain itself is larger than any course, which is limited to a finite period of time; thus, multiple courses, at different levels of expertise, are common. Note also that there may not be good coverage of the domain by any particular course, nor do many courses include practice that leads to real mastery of the tasks that domain experts perform. This possible disconnect between the course and the domain represents one need a mastery system might fill. Another is to mark progress by providing milestones with performance standards so that the learners' progress along the way may be demonstrated and appropriate actions taken.

When embedded within a curriculum, or more properly, within a specific course, a mastery assessment system has two important functions in relation to the curriculum:

- It gives trainees and supervisors a view of the domain to be mastered that may be broader than a particular course. It does this with a "mastery map." In this way, it helps establish the representativeness and appropriateness to the domain (as known by experts) of the elements selected to be mastered in a particular course.
- It enables instructors and learners to determine the levels and standards of achievement along many possible paths to mastery. These are communicated not by numerical abstractions, but by anchors and reference tasks that refer to different levels of growth and to achievements in the world of work.

Development of a mastery assessment system involves an interdisciplinary mix of measurement, cognitive, and instructional sciences. Because the domain may be broader than any particular course, the mastery map must be customized in a manner analogous to an instructor selecting chapters from textbooks.

Mastery beyond minimum competence

When a measurement organization obtains group consensus on learning milestones in a particular subject area, that consensus usually converges on what might be called a minimum competence standard. The mastery that is assessed in a mastery system, on the other hand, looks forward to a time when, after long commitment and effort, the learner has obtained a life-long capability. It is important therefore in developing mastery assessment systems to identify people who are truly masters or exemplars of the performance to be achieved by the end of the training.

Mastery signifies achievement of performance goals of a high order. Mastery is personal and unique and is achieved after long periods of persistence and commitment. Performance engineering can identify the potentials for improvement in well-defined tasks and teach these performances to others. However, the United States needs workers capable of high levels of flexible performance. The term mastery reflects this need by going beyond exemplary performance in a single, well-prescribed job. Training programs must aim toward development of mastery because jobs change rapidly, and workers must be prepared to adapt to changing demands.

The assessment of higher levels of mastery often involves examining unique productions (complex problem solving, oral presentations, written analyses, portfolios, etc.). Some precursors to mastery can be assessed at earlier levels of growth by encouraging the learner to practice some element of mastery appropriate to the growth stage. At intermediate stages of learning, the trainee can experience the rewards of persistence and the ability to expand upon what has been learned until able to do something extremely well.

Assessment includes the use of standardized measures of competence and guidance to judge the precursors of mastery at various levels. These measures may include disciplined subjective scoring or, in the future, intelligent computerized scoring. Instructionally sensitive assessment will have new properties and paradigms that are not yet fully developed by a measurement science built to support certification, selection, and classification.

Components of a mastery system

The most effective and complete examples of mastery assessment systems will be implemented in heavily computerized learning assistance centers or in work areas with local networks. Simpler forms of mastery assessment systems that could be implemented partly on paper also could be developed, as well as transition technologies that start with paper delivery and eventuate with the heavily computerized model presented in this chapter. For the simplest system, at least one computer should be available for scoring and record keeping. Such a computer provides a primitive learning information system.

Major non-hardware components of a mastery system should include the following:

- a mastery map usable by learners and instructors to envision and communicate learning goals;
- reference tasks;
- calibration of items and reference tasks;
- an instruction-oriented scoring system for each reference task;
- a professional development program to help instructors/supervisors learn to use the system effectively.

These components are intended to serve instruction and to be linked with instructional components. Instructional components will include repeated practice in reference tasks, subscoring to guide the instructor in coaching, and report-generating systems for both learners and instructors. Coaching, in this context, is analogous to the instructional process an excellent athletic coach uses. The process may include modeling the desired performance, observing practice trials, prompting, encouraging, and fading the prompts as the performance becomes adequate.

The mastery map. Each training enterprise requires that a domain be mapped and defined in terms of knowledge elements, job elements, and the kinds of tasks that novices, adepts, and masters perform. The mastery map is a way of making the domain visible. The mastery map gives the trainees and their supervisors or instructors an overview of the journey at the beginning of learning. The mastery map also permits communication about initial placement and next steps for each individual trainee. The mastery map could be displayed on a large wall for all, but individual maps with status information also would be made available graphically on computer screens, hierarchically organized and shown a screen at a time.

A key feature of the mastery map is information about the current status of the learner or trainee. One way to mark progress is to color sections of the map red, yellow or green, depending on whether a person has passed through a challenge successfully (green), needs further coaching (yellow), or has encountered such difficulty as to indicate a lack of readiness for the challenge (red).

There is no implication here that each challenge needs to be passed. Different paths are appropriate for different individuals. There is no demand for passing tasks in a particular order. A mastery assessment system is a flexible use of measurement to aid learning and instruction. The term "mastery learning" sometimes describes instruction that demands rigid adherence to a linear sequence of tests of minimum competence, with no choice but to pass each test before going on. The mastery system concept presented in this chapter promotes a more expansive (and more historically accurate) view of mastery. It describes a human master as one who has attained exceptional levels of competence, not the narrow definition implied by "mastering a test." This system also offers different approaches to delivery and implementation.

Reference tasks. A reference task generally is more complex than a single item. It may be a testlet (a related group of items) (Wainer & Kiely, 1987), a curriculum-embedded exercise requiring multiple responses, or a simulation exercise. A reference task is contextualized. It refers to some real-world work that communicates to intelligent lay people the relevance of what is being practiced. The scoring of a reference task also may refer to component process constructs important to exemplary performance on the task and useful in coaching. A record of an individual's accomplishment on reference tasks can serve to build self-confidence. In this way, it may serve as a reference example—a benchmark to look back to when attempting new challenges. Table 1 contrasts test items and reference tasks.

Calibration. Reference tasks can be placed on scales to show the degree of growth they represent. Test items, perhaps grouped into clusters or testlets, can also be placed on such scales. For example, the following tasks were used in the NAEP study of the literacy of young adults (Kirsch & Jungeblut, 1986). They assess literacy skills used in interpreting documents (forms, maps, charts). The scale values are statistically determined measures of difficulty based on the item response theory. In the literacy study, they are described and explained in terms of task features that account for variations in difficulty.

Table 1—Test items vs. reference tasks

Test Items	Reference Tasks
Usually administered via paper and pencil	Usually administered via interactive computer
Written objectives prescribe test items	Flow-charts/interaction specifications prescribe reference tasks
Single response, usually multiple choice	Multiple responses, which together provide for a quantitative assessment of degree of success
Dichotomous scoring: • Pass • Fail	Trichotomous scoring: • Pass (competence demonstrated) • Needs coaching and practice • Not ready for this task
A complete test with subtests can be used for diagnostic purposes	Simultaneous subscores, taken to measure component processes or states, provide data to guide coaching
Items and entire tests are often decontextualized. (Lay people may not see the relevance of the questions to valued capabilities in the real world)	Tasks refer to or simulate aspects of valued real-world activity (e.g., in a job)
Items can be calibrated and placed on a measurement scale	Reference tasks are related to one or more statistically validated scales that span the domain; calibration of reference tasks is a current challenge for measurement science
Except in CAT systems, administration of the next item is fixed by its order on the page	Order of requests is determined dynamically
Test items are "used up" after one attempt and are of little value for repeated practice	Although some reference tasks (e.g., paragraphs to read) require files of alternative stimuli for practice, many (e.g., simulations or game-like events) can be practiced repeatedly without using up material
The objectives and specifications for developing the test items are not suitable for learners to view and are not presented to the learner	A model of mastery can be made available to help learners observe the attributes of successful performance

The following examples illustrate the calibration of reference tasks in a mastery system:

- *Sign your name on the line that reads "signature."*

  Scale value: 110

- *Put an 'x' on the map where two particular streets intersect.*

  Scale value: 249

- *Fill in a check to pay a particular credit card bill.*

  Scale value: 259

- *Use a bus schedule to answer: "On Saturday morning, what time does the second bus arrive at the downtown terminal?"*

  Scale value: 334

- *Use a bus schedule to answer: "On Saturday afternoon, if you miss the 2:55 bus leaving Hancock and Buena Ventura going to Flintridge and Academy, how long will you have to wait for the next bus?"*

  Scale value: 365

Initial calibrations could reflect both the experts' judgment and early empirical results. Later calibrations with large numbers of learners would be based on a new generation of statistical models appropriate to that end and would become more accurate and stable. Once the reference tasks in a set are calibrated, the scale values gain meaning by showing the constructs of knowledge and skill required to succeed at tasks with a given range of values. Each mastery system has its own calibration. The scale and scale interpretation are developed and validated for a particular content, level, and purpose.

Instruction-related scoring of reference tasks. Test items are normally scored dichotomously—correct or incorrect. It is possible to score reference tasks more finely to connect performance with instructional strategy. One important aspect of instruction-related scoring is to identify common or especially troublesome misconceptions and erroneous procedures and to provide appropriate feedback.

Reference tasks may be very close to the criterion of real-world performance. These reference tasks offer a means to measure critical dimensions of performance, such as speed, accuracy, and strategy. In this way, the reference task moves closer to being the actual criterion than to being a predictor of later performance. Sometimes computer simulations are "better than the real thing" because we cannot measure the real thing in situations that are, for example, extremely hazardous or costly, or that pose other formidable difficulties.

A professional development program for supervisors and instructors. A mastery system is always accompanied by a professional development program for supervisors, who often are also the instructors. Professional development includes training in the appropriate use and interpretation of measures and in methods to build and maintain appropriate climates for teaching and learning. The learning information system provides a tracking system so that instructors can make professional decisions about how to manage the paths of a group and its individual members. More advanced versions will provide information to guide the coaching of individuals and groups. Instructors will have to learn to use these technological tools to be successful at using the mastery system for placement, for the various aspects of tracking, for different aspects of training management, and, ultimately, for diagnosing and coaching individuals and small groups.

A mastery system is designed to function within a cooperative community of learners whose goals are to help and encourage one another, to teach one another, and to facilitate the maximum amount of learning for all. Users must learn to build and sustain such environments.

How Can a Technology-Extended Measurement Science Benefit Training and Performance?

Some possible answers to this question can be seen in the framework of Instructional Systems Development (ISD). The steps in ISD are frequently divided into five categories (see Branson & Grow, 1987):

1. front-end analysis;
2. design activities;
3. development activities;
4. implementation;
5. continuing evaluation.

In this part of the chapter, the section headings correspond to all but stage 2 of these five stages of ISD: front-end analysis, instructional development (featuring prototypes that use data for iterative improvement, implementation, and continuing evaluation (emphasizing ongoing data collection for continually controlling and improving the interactive training system). Except for the impact of construct valid scales to represent the domain and construct-related scoring within reference tasks, there is no further discussion of possible impacts of measurement science on design activities. Before going into these stages, we will describe a scenario of future training application in order to illustrate how training can be enriched by technology and measurement science.

A Description of a Future Training Application

The following scenario describes how a new delivery system involving interactive technology with integrated training and measurement can be installed in a company to solve a large-scale training problem. The scenario also includes an example of a transition plan for moving from a current training environment to a new technology- and measurement-intensive environment.

The application

Assume that a large company operating at numerous locations has decided to install computer work stations, networked into clusters for functional work groups of 10 or more workers. The software to be implemented includes some fairly complicated mail-order data processing software, which is being introduced slowly to replace manual procedures. This new software is integrated with word processing, spreadsheets, and other productivity software packages so that clerical workers have a clear career path. The goal is for many trainees to start at the bottom, learning basic and repetitive aspects of the mail-order system, and for some of them to move up to more advanced secretarial and administrative positions. The company has experienced substantial turnover at the secretarial and clerical levels.

The talent pool

Because of turnover and because the available labor pool is not well-prepared, the company finds it must dip more deeply into the less educated levels of the labor pool, where there are literacy problems. The available candidates are not illiterate, but their literacy skills are frequently insufficient for keeping up with the company's training classes. In addition, the company's proposed software for automating the manual mail-order processes has proven to be extremely training-critical. That is, exemplary performers can obtain substantial increases in productivity, but when those having lower literacy skills or without proper training use the automated system, there are numerous costly mistakes. The company is not anxious to implement automation widely or rapidly until the implementation can progress without these mistakes. In this scenario, the company has begun installing the clusters of microcomputers in a few of its most exemplary work groups and has made some initial attempts at training the old and new workers.

The current training activity

The personnel department has established a series of classes and a "lab room" with machines installed for practice in using the software packages. It also has an agreement with local community colleges to provide literacy classes where needed. The process of training in the lab room is low-key, as there are not enough trainers to expand it to the workforce at large, nor is the laboratory adequate to support many trainees. Management would like to decentralize the training activity to the locations where the machines will be installed. But because of the current training constraints, management has slowed the rate of installation of machines.

Management's plan for introducing distributed measurement-intensive training

Management's plan has several steps:
1. conducting an intensive front-end analysis using some new concepts for defining and placing a value on expertise;
2. developing a mastery map, reference tasks, and scoring procedures based on the front-end analysis;
3. instrumenting the existing classes for continuing measurement;
4. selecting and adapting existing interactive training materials and developing new materials;
5. iterative testing and revision at a small number of test sites;
6. implementation plans and tactics for replication to multiple sites.

Conducting an Intensive Front-End Analysis

The company managers first want to make clear the level of expertise they hope to achieve. They want both themselves and the project personnel, including the trainees, to see a visual representation of the mastery to be obtained and to be able to mark individual and group progress visibly, through status information on a mastery map. In order to mark trainee progress, they want reference tasks to be identified and measures developed to assess performance on all reference tasks during the course toward mastery.

Management commissions a front-end analysis that will yield these products, as well as the other products of a more conventional front-end analysis. As in the past, they expect the analysis to yield job tasks and objective performance levels, but they want more than verbal descriptions. They want samples of the products and data on the speed and accuracy with which work products can and should be produced (work standards). They want diagrams of the procedures followed by more expert workers. They want a list of the names of the most expert workers who have been located, within or outside the company, so that these people can both provide data on exemplary performance and serve as models.

Management also wants an economic analysis of the payoff from a target level of expertise. They wish to locate or develop a quantitative model demonstrating the benefits of expertise and the costs of mistakes, multiple revisions, and slowness due to lack of expertise. They know that as one alternative they can commission a performance audit, following the methods outlined by Thomas Gilbert (1978). A performance audit yields the ratio of the performance of exemplary performers on tasks and subtasks to the performance of typical, merely adept, performers. These ratios are the Potentials for Improving Performance (PIP). The performance audit identifies which PIPs have the greatest potential and how these are related to costs and economic benefits. A team is commissioned to seek this type of economic information from vendors or, if the information is not available, to recommend an action plan and a budget by which the company may pursue these questions to the depth desired.

Another aspect of the front-end analysis is to determine the beginning performance levels of individuals from the labor pool, including the ranges of literacy available and productivity and the minimum keyboarding and computer use skills. A "job literacy" analysis is essential to determine the levels required for success in the job training and on the job. Consistent with the needs and with the capabilities available in the pool, analysts seek to determine the literacy standards needed to reach exemplary job performance levels. Management may then commission appropriate literacy instruction so that more applicants from the available pool can benefit from job training.

Developing the Mastery Map and the Reference Tasks

After the front-end analysis is available, the project team develops the initial version of the mastery map and a set of (initially sparse) reference tasks. Fortunately, the team finds a commercial package, which provides speed and error scores with norms for the word processing and related keyboarding skills. The team commissions the company producing this package to produce assessments based on reference tasks to define increasing levels of expertise in the use of the mail-order processing software. This company has authoring software appropriate to keyboard oriented jobs and agrees to produce rapidly the new assessment tasks for the order-processing software.

Instrumenting the Existing Classes for Continuing Measurement

The performance measurement systems are installed in the workstation laboratory so that at certain points during practice and learning, as well as at the end point, the proficiency of the trainees can be measured. The wall charts of the mastery maps are hung in both classroom and laboratory, and the initial computerized mastery maps are installed on the computers.

Paper-and-pencil tests and other group-administered measurement activities also are introduced into the current classes to help get a better assessment of how the trainees are progressing from day to day. It is planned that this instrumentation in the laboratory and the classroom will produce data that can be used to improve both of these training resources and also to guide the implementation of the distributed training environment.

Selecting and Adapting Training Materials

The developers look for available interactive training material so that the training can be decentralized from the training room to the work units. They are able to find some commercial packages for the office productivity software, although the packages require some new development of reference tasks, as identified on the mastery map but not covered by the commercial lessons. The developers use "lean development methods" by producing things rapidly for presentation by the instructors in the training room and putting on the computer only those materials that have been tried out with the trainees and have shown their ability to communicate clearly. The new reference tasks, initially used only for measurement, provide files of response data from which lists of common errors are obtained. Feedback messages are prepared and reference tasks are gradually improved by providing better and more thorough feedback. It is planned that the training will be administered largely interactively once the system is more fully developed. As an interactive, on-demand facility, it will continue to be used after implementation, under the direction of the supervisors, as turnover brings in new workers.

Iterative Testing and Revision

While the instrumentation for iterative testing is being installed, a schedule is worked out for implementing the initial pilot-test work stations. The pilot-test work groups are identified, and the equipment is ordered for these people. A training schedule is established combining training room activities with practice at individual work stations using real work assignments. The general strategy is to use the networked computer work stations in addition to the practice equipment in the training room. Thus, the development of the second stage of the implementation plan builds on what is being learned in the practice lab and transfers into a plan for the work units.

Time for extensive revision is scheduled between each of the installations of the software and training materials into a pilot-test work group. In this manner, it is hoped that the interactive practice materials, the on-line measurements, and the implementation plan can all be improved.

Great care is taken during the pilot testing to develop and refine an implementation plan. The team finds that the supervisor is the key to successful implementation. The supervisors are brought to one of the training classes and are taught how to tutor individual workers in the practice laboratory. Using the identified expert workers is also found to be an extremely effective tactic in the implementation process. The expert workers' speed and error statistics are made available, and they come to the pilot-test sites to talk to the workers, answer their questions, and encourage them. As new workers achieve high levels of proficiency, these new "masters" begin to coach and encourage others, both in their own work groups and beyond.

Implementation Throughout the Company

The implementation strategy allowed for three iterations in pilot-test work sites before the full-scale implementation was conducted throughout the company. Not all reference tasks were available until the third pilot test, so each iteration involved a more complete system. Not only was there a desire to make sure that the training materials and interactive measurement materials were soundly conceived, but there was concern that all the human management elements of the implementation of this technology be explored prior to full-scale implementation.

A description of the mature system

The company management found that it took 16 months to go through the development and pilot testing and another six months to complete the full-scale implementation. During this process, reports provided continual feedback on the levels of mastery that the trainees were achieving, made possible by the instrumentation installed in each work group. The reports were used by line supervisors and were well understood by the top management group. The economic analyses for the program convinced management of the value of paying full attention to the development of human expertise. Models of the cost of not having the needed worker expertise thoroughly justified this investment in human resource development. The models were refined continually during the implementation, and the performance measures were revised twice a year to make sure that they were current and had not become familiar to trainees.

Management was very pleased with the successful implementation of this project, but they knew that the job was not finished. Not only had they not reached the top levels of exemplary expertise defined before the project began, but technology had moved ahead during the 22 months of development and implementation. It was possible to identify individuals who had broken through the ceilings defined by mastery levels established in the front-end analysis. These "star performers" had exceeded the previous standards for exemplary performance, using a combination of improved software productivity tools and higher levels of human expertise.

Management wished to continue the growth that had occurred by setting clear standards for the growth of mastery and by having a measurement system to monitor it. But they did not wish to make the tactical error of measuring individual performances in a hidden or manipulative manner. Therefore, they requested only aggregate data from each work group, and they established quality circles so the workers within the work group could study the individual data themselves and try to raise their own performance levels. In so doing, each work group (including the supervisor) was rewarded both for increases in productivity and for providing suggestions and improvements of use in the total human development system.

By implementing a continuing evaluation and improvement cycle and delegating it to the workers involved, along with a proper reward system so that the workers could share in the benefits of the productivity gains, management was able to observe a continuing upward trend in the productivity and expertise statistics. When work groups made requests for additional hardware and software, management had a framework for evaluating the economic tradeoffs of installing it and for providing the on-the-job training needed.

Summary

In this scenario, several features of the measurement- and technology-intensive approach being proposed can be singled out.

- Management used a front-end analysis that clearly identified high levels of human expertise, based on real people, not just standardized descriptions of job elements and tasks.
- Economic costs and benefits were a part of the front-end analysis; there was a serious attempt to relate human expertise to its economic payoff and lack of expertise to its consequential costs.
- Management did not develop a one-shot plan a priori, announce it with fanfare, and then distract the whole company with a crash program to automate. They moved slowly, reducing their costs and risks at each step. They first introduced change to a few pilot-test sites and identified problems through the use of test data. They depended on good training developers to solve the problems identified

by testing while the problems were small and contained. This unusually enlightened (and admittedly fictitious) management group did not expect perfection the first time, but planned for problems in implementation. They provided the time and means to correct these problems.

- The management group took direct control of the human development process and used its training and development resources as a line rather than staff function in implementing productivity improvement through technology and human resource development.
- The management of implementation was seen to be more important than the development of effective training, which, in turn, was seen to be more important than the purchase of appropriate hardware.
- Good management of human resources through the continuing collection of performance data was not seen as a tactic only for the implementation of an initial change, but was regarded as a tool for continuing improvement.
- The use of data to improve performance was delegated to the workers. A climate of trust and growth in expertise and productivity was pursued for the good of both the company and the individual workers.

Multidisciplinary Approaches to Front-End Analysis

In the scenario, the idealized management group employed a kind of front-end analysis that is an important emerging multidisciplinary technology. This type of front-end analysis cannot be described solely in terms of traditional job and task analyses, although these tested methods (McCormick, 1979; Fine & Wiley, 1971) are important and valuable tools.

Front-end analysis: A problem in representation

We owe to our colleagues in the field of artificial intelligence a better understanding of the power of different representations in shifting information processing from something extremely difficult to something easy—sometimes trivially easy. Selecting the wrong or the right representation for organizing the data and symbolizing the important relationships for a computing problem makes the difference. Bruner (1968), in his essay on instructional theory, argued that the idea of different ways of representing information to learners can make a very large difference in the difficulty of learning or solving problems. A basic concept in the selection of different representations is that their economy and power is maximized only in relation to a given *purpose*. Different purposes determine which representations will be awkward and which economical and powerful.

The usual product of front-end analysis—lists of tasks, knowledge areas, and objectives—is used by training developers to produce a syllabus and to guide the design of instructional materials. It has been used by test makers to produce specifications for test items. Its purpose has *not* been to produce a top-level guide to the structure and content of a domain of expertise. Such a guide would maximally communicate about the domain to learners and instructors and would provide a way of tracking progress from novice to higher levels. Both mastery maps and reference tasks as defined in Part I can serve as representations suited to these trainee-centered purposes. The revitalized and broadened form of front-end analysis we seek will sensitively select different representational forms for the different purposes of a front-end analysis.

Currently, the job tasks and knowledge areas are represented to the analyst and to the questionnaire respondents alike in terms of written words. This form of representation is best for describing the *knowledge areas* of work. Procedures can be described with words, but they are then described at a knowledge level, which is not the same as the tacit expertise of a true master, who possesses it "in the fingers and in the gut" and is able to perform each procedure effortlessly at the proper time. These tacit components of job knowledge are often associated with the highest level of expertise, which has become automatized in the human expert. Tacit components usually are missed when the representation is solely verbal. Polanyi (1962) provides a discussion of tacit knowledge. New methods of modeling expertise should be investigated. By modeling we mean making a visible representation of otherwise invisible expertise. Words and numbers are not the only form of representation. Some of these new methods are being pioneered through that branch of the field of applied artificial intelligence called "knowledge acquisition," the most difficult part of knowledge engineering (Hart, 1986; Bunderson, 1987). Knowledge acquisition means acquiring knowledge from one or more human experts and representing it in a publicly visible and usable form, usually symbolic.

A knowledge engineer uses other forms of representation besides words and emphasizes problem solving, not just declarative knowledge (terms, facts, categories). In this way, knowledge engineers penetrate to a higher level of expert knowledge. A knowledge engineer thinks in terms of a knowledge base of formal rules, a data base of facts, and a set of informal heuristics or rules of thumb that experts use to guide the use of rules and facts. (A good heuristic is a guideline that tells the expert where to search for a solution and where *not* to spend time or effort.) The rules are stated in a formal language such as predicate logic, which is more formal than standard written English. Despite this, if left at the level of a representation on paper, such knowledge is still only a symbolic formalization.

It is an unusual formalization, however, because it can be tested. The rules in the knowledge base and the facts in the data base may be exercised by programming a prototype expert system and testing it by submitting cases and problems to see if it will solve them as effectively as experts solve them. Producing a complete expert system is obviously too costly for the ordinary front-end analysis, but many problems in knowledge specification can be discovered early through attempts at formalization and through the discipline of testing the formalization (manually, without a computer) using sample cases that the expert is supposed to be able to handle. Consideration of work behaviors to this depth (that is, to the level of the rules and heuristics that experts use in doing their work) is a discipline which has not yet been reduced to practice in job analysis.

In testing an expert system, the selection of test cases is of utmost importance. By analogy, the selection of reference tasks becomes absolutely critical. These reference tasks become the focus for the development of practice tasks to be used during training and proficiency testing.

Reference tasks

Like the test cases for assessing the quality of an expert system, a good set of reference tasks will span the range of high-level expertise. These tasks define the upper reaches of the domain. Unlike test cases in an expert system, however, reference tasks must also be developed that span the lower ranges of expertise—novice, advanced beginner, intermediate, adept, etc. These levels of reference tasks are needed in a complete mastery map that tracks the trainee from lower to higher levels of expertise. Selecting a set of reference tasks is an exercise in assessing the representativeness, completeness, and appropriateness of the domain coverage. It is one aspect of the validation process for an assessment and instructional system.

Challenges to measurement science

There are many challenges to educational and performance measurement arising from the need and opportunity to improve education and training dramatically. The international economic challenge makes it vital that we succeed in this interdisciplinary endeavor. The challenges to measurement science can be considered as different aspects of the processes of establishing validity.

- Validating the representations of the domain: Are the mastery map and the lists of job tasks and knowledge areas representative and complete? Does the mastery map communicate clearly?
- Validating the scale constructs derived from the analysis of the tasks in the domain: How many scale dimensions are sufficient to characterize the domain at the top level of the mastery map?

- Validating the construct-oriented scoring of different reference tasks on different scales: Do scores or states achieved in a simulation task reflect actual expertise in the simulated job?
- Validating the cognitive and instructional constructs: Do subscores that reflect cognitive constructs, when used in instructional feedback and advice, lead to performance gains?
- Validating the implementation plan: Does it work? How can it be improved? More generally, what principles of implementation have cross-situational applicability?

Using reference tasks to measure growth raises some immediate challenges to measurement science. In this area, measurement science must stretch itself in new directions in order to rise to the opportunity of creating instructionally sensitive assessment. Unfortunately, much of the measurement work developed with test items, which usually are presented at the beginning or end of training, is not applicable to the problems of measuring growth "in process" using reference tasks.

Individual test items are ideal for quickly sampling large domains of knowledge: facts, terms, concepts, and simple applications. But short, single items are not well adapted for many kinds of job elements. For example, multiple-choice items must include the right answer, which may be recognized or found by elimination, rather than being recalled or generated. Interactive computer systems can simulate many kinds of job elements not well-assessed by short, objective items. They provide a way of administering the reference tasks that introduce much more of the job context into the training and testing situation. Researchers are attempting to define this building block of measurement, the reference task, and to find how it is different from single test items or clusters of test items.

Cognitive science is an important element of this enterprise because cognitive science helps tell us what is important to measure. Instructional science helps us use the scores to provide informative feedback that guides the students in reorganizing their knowledge and skills so that they can perform at the next higher level of expertise. As cognitive and instructional science produce a new set of constructs, measurement scientists then are challenged to model these constructs through scores that can be shown to be construct valid. That is, variations in the scores reflect variations in the underlying constructs.

These new constructs must be instructionally sensitive. Older measurement constructs, such as aptitudes, abilities, and generalized traits, have not proven to be useful in guiding instructional activities. Studies of aptitude by treatment interactions have shown some individual traits that would lead to prescriptions of alternate instructional treatments, but in general, the constructs of greatest promise for providing instructional feedback are the cognitive and instructionally oriented constructs that must be derived from deeper analyses of specific tasks. In time, measurement scientists will be challenged to model constructs that cut across a variety of tasks.

Measurement science is being severely tested by these demands of the real requirements of training. To develop scoring procedures using reference tasks instead of individual test items requires a new look at standard measurement models. To array the reference tasks on a mastery map with a scale of proficiency going from novice to expert requires new foundations for the calibration of tasks. In existing measurement science, there are simplifying assumptions that the ability being measured is fixed, rather than changing, and that one task has no influence on another. The opposite assumptions are made in training: Ability is changing as rapidly as possible as a consequence of practice on the tasks, and one task transfers positively to another. Measurement scientists generally have been only marginally familiar with the concepts and methods of training, instructional systems development, and instructional science. Broadening the contributions of measurement to training requires that measurement professionals learn of the problems and successful principles in the training field. The payoff for this learning may be great, because mastery maps with calibrated reference tasks that can provide continuous measurement both during training and afterward can potentially reduce costs and improve productivity enough to more than justify the effort.

Tacit knowledge

Knowledge engineering is only one approach to making expertise visible. It still produces written formalizations even though they are written in a more precise and formal code implementable on a computer. True expertise is not limited to knowledge of facts, procedures, rules, and heuristics. It includes tacit components and human motivational components. One example is "role commitment," the commitment to exert effort toward a new role—to practice over the long periods of time necessary to truly master an area. Role commitment implies, among other things, persistence of effort.

There are other subtle human dimensions in the work environment. It is important to include real people in a front-end analysis and to use them to model and motivate during the implementation and training process. Bandura (1977) has developed a theory of social learning that uses human models as a central element. The human models carry and communicate tacit knowledge and role commitment. As new people enter the elite group of experts, they can be added to the training resource pool and honored, as a reward for their effort and as an incentive to others. Also, because much of the expert's performance is never reduced to written or symbolic formalizations, human experts can often notice ways that the trainees can improve and can coach them when given a chance to observe and do so.

Summary: Toward advances in front-end analysis

- The practice of job and task analysis is being expanded by multidisciplinary approaches that introduce methods from knowledge engineering, cognitive science, and the social/behavioral sciences.
- Measurement methods using sampling and data collection can enhance the relevance and completeness of front-end analyses and its sensitivity to subgroup differences.
- Mastery maps are a suggested new product of front-end analysis. They fill the purpose of modeling the domain, organizing the reference tasks, and tracking movement through the domain for the benefit of trainees and instructors.
- Analysis methods must be expanded beyond representations that use only written descriptions of work behaviors. Also important are formal representations of critical facts, rules, and heuristics. Tacit human elements are vital but often can be conveyed only from one person to another through a modeling, coaching process.
- The selection of appropriate reference tasks at different levels of proficiency is a critical part of front-end analysis. The reference tasks refer to real-world jobs, and to the simpler tasks that novices, intermediates, etc., can perform. They are contextualized, unlike most test items or individual knowledge topics.
- A revitalized front-end analysis can lead to more informed development of needed training materials. The domain specification leads to a mastery map. The reference tasks lead to practice exercises and are the building blocks for providing both proficiency testing and individualized feedback.

Using Data to Improve Instructional System Development

Bootstrapping: A lean development approach

The scenario described a "lean development approach," meaning that the new instructional system was bootstrapped by building on and adding incrementally to an old one. The scenario did not provide all of the details for accomplishing this. Some of the methods used were introduced in the last section. Front-end analysis can be used to produce a mastery map that represents the domain to be mastered and provides a description of the knowledge areas and reference tasks to be dealt with at the different stages of development from novice to expert. Imposing a mastery map on the existing training can reveal where the training is deficient in its coverage. Missing elements may be added either through interactive lessons on the computers themselves, through lessons taught by the instructor in a stand-up mode, or through printed texts.

The mastery map can serve as a management structure for developing and testing lessons as well as for guiding trainees and instructors. As reference tasks are developed they are checked off. The training situation provides a lean approximation of the integration of tasks and lessons into one course of study. It is "lean" because most of the context is provided by the instructor, who also provides feedback and guidance for the slower learners. Interactive lessons have to stand alone, aided only by the supervisor or fellow workers at the work site. They must carry the needed context and provide sufficient feedback for the slower learners. In lean development, the training center is used and measurements are taken from it during the time the interactive lessons suitable for distribution to the other sites are being located or developed.

During their development, the reference tasks may be used to enrich the instruction. Later they can be used in the distributed locations. An ideal reference task is developed in such a form that it allows measurement to occur but provides for repeated practice. In word processing, for example, a reference task may consist of typing a certain kind of document using certain word processing features. It is easy to generate new practice tasks for this kind of exercise. The trick is to provide the scoring so that proficiency can be made visible to the trainees and instructors in terms of some overall scale, so that the scale relates to norms for a larger group, and so that individual feedback messages can be generated.

As these computer interactive reference tasks with built-in scoring are introduced into the training, they enrich the ongoing classes and bring closer the day for distribution to remote sites.

Smoothing the path up the mastery mountain using difficulty statistics

If trainees are given a task too easy or too difficult for the their current level of progression, practice is less than optimal and may even be detrimental. It may prove to be either boring or discouraging. Therefore, it is important to know the relative difficulty of reference tasks. Which reference tasks should be presented to novices, to intermediates, and so forth? Measurement can aid in this process by developing models that use data to assess the relative difficulty of each reference task on some underlying proficiency scale or scales.

Very often this empirical process can reveal "cliffs of difficulty" in the sequence of practice tasks. These are the points at which trainees begin to have difficulty, fail to make continuing progress, and encounter frustration. Providing time for iteration, as was suggested in the scenario, allows developers either to present their reference tasks in a different, more accessible manner, or to build a set of scaffolding exercises that can lead the trainee more slowly up the cliff of difficulty. Sometimes tasks are presented in an inefficient and ineffective sequence that creates unnecessary slopes of extraordinary difficulty for many. The use of rough difficulty statistics can reveal this condition and can lead to suggestions for better sequences of reference task exercises, based on a clearer understanding of what is going on at the difficult zone.

Cognitive science may reveal that a cliff of difficulty reflects the need for an underlying reorganization of cognitive structures. Attaining expertise is not a smooth, continuous process that grows task by task. Higher levels of expertise are often reached only by restructuring the knowledge, skills, or procedures of a domain in some important ways. Cognitive science provides tools for investigating the constructs involved in the reference tasks found around cliffs of difficulty. Complexity in underlying structures makes it difficult to scale and calibrate reference tasks at these ranges of discontinuity.

Cognitive science uncovers the structure of knowledge and expertise and the restructuring that successful learners generally accomplish. This may make more precise measurement models possible, while rough initial measurement could be used to focus the costly cognitive analysis in the important transition zones. Instructional science provides guidance for developing effective training presentations and informative feedback methods that will be optimally effective in guiding trainees through these difficult paths toward restructured knowledge.

Neither measurement science nor the disciplines that have contributed to training systems development have converged on a standard set of methods for scoring and calibrating reference tasks so that the advantages discussed in this section can be achieved. Much cooperation between cognitive scientists and measurement experts is necessary before this will be possible. For example, measurement science is still in the process of developing good models for scoring reference tasks based on underlying cognitive structures and processes. Cognitive science lacks good case studies of the cognitive analysis of restructuring that must occur at the points of discontinuity.

Summary: Impact of measurement perspectives on instructional development

Using the mastery map as a management structure and using iterative testing and revision in a development approach are not especially new ideas. They represent a productive interplay between good measurement and iterative instructional improvement. They are used less frequently than is warranted by their potential payoff. Training development practice can further be enhanced by emerging contributions from measurement science, including determining the relative difficulty of reference tasks and providing the construct-related scoring of such tasks, tied to instructionally effective feedback.

Toward a Design Science of Implementation

The scenario conveyed a carefully planned process of implementing change slowly and iteratively throughout a company. It suggested that the implementation process be directed by line management as a fundamental responsibility for the management of change. The process should involve measurement to test each step of the way and to validate an implementation plan before extending it to the whole company.

Anyone familiar with the attempt to improve productivity, whether through training or job aids, is well aware that the majority of the variance in success or failure is attributable to implementation. The best job aids, the best training plan, can fail if implemented improperly. Implementation, and the design of implementation plans, strategies, and tactics for given situations, is worthy of being called a new interdiscipline in its own right.

The term *design science* used in the heading for this section is based on the usage of Simon (1981), who in his provocative and prophetic book *The Sciences of the Artificial* defined a science of design as

> a body of intellectually tough, analytic, partly formalizable, partly empirical, teachable doctrine about the design process....
>
> Such a science of design not only is possible but is actually emerging at the present time. It has already begun to penetrate the engineering schools, particularly through programs in computer science and "systems engineering," and business schools through management science.... We can already see enough of its shape to predict some of the important ways in which engineering schools of tomorrow will differ from departments of physics, and business schools from departments of economics and psychology (p. 132).

The search for sets of principles to guide the process of design is general to all stages of Instructional Systems Development, not just the design stage. Prescriptive principles can guide front-end analysis in seeking and collecting those representations of knowledge that will serve a variety of design and development purposes. Principles will help establish the content and construct validity of the domain representation and will guide the designing of mastery maps, the selection of reference tasks, and the development of feedback messages.

The implementation process has not usually been seen as a scientific activity. It has generally been viewed as a practical concern of managers. The high frequency of failures in implementing new systems and programs with significant potential indicates that this view may be mistaken. Research-based principles for designing implementation plans and carrying them out must be validated and shown to work and to produce results. The principles we seek are clearly goal dependent; hence they are better sought within the framework of a design science than a descriptive science. Designs always involve human purposes.

At Educational Testing Service, we have organized a new group of researchers in the Division of Educational Policy Research. We call this group the Implementation Policy Group. Their job is to investigate implementations of new instructional and assessment systems and to seek general prescriptive principles for introducing technology-intensive applications similar to the one described in the scenario. They seek to discover a set of generalizable principles that can be used to guide the preparation of implementation plans, strategies, and tactics.

A set of principles for guiding the design of an implementation plan will include contributions from a variety of fields. Engineering and measurement are two of the fields that can contribute to the development of automated performance measurement systems and systems for formative purposes. Lean development approaches that provide a bridging strategy from current practice to new practice will also be important. The principles of organizational behavior and other social sciences are needed for defining methods to change the goals, roles, and technological tools of work groups as technological change is introduced.

In the scenario, the supervisors had to adopt new instructional roles, but new learning and working roles were offered to the workers as well. First, a longer career ladder was introduced, extending from the lower clerical levels to the higher secretarial and administrative assistant levels. Also, workers and supervisors were given the challenge of improving productivity themselves at the end of the training implementation.

One of the most important principles of implementation policy will be the exercise of patience. Goals, roles, and traditions do not change readily. Good management, good instruction, good modeling on the part of those who are committed and can demonstrate success—all these are necessary to change the human parts of a culture into which new technology is being introduced. Naisbitt (1984) had a term for it—"high touch." As he pointed out, high tech will be rejected unless there is a large dose of high touch—the human element—to smooth the transition.

Continuing Evaluation

Summative evaluation is the most common concept of evaluation in the minds of policy makers. Moreover, the policy makers in a company are happy to forego any kind of evaluation unless their opponents demand "scientific proof" that a certain program is a success or that a certain program should be dismantled or replaced.

Through the concepts of calibrated scales, with reference tasks as anchors, measurement science can provide evidence to assist policy makers in resolving legitimate concerns about the effectiveness of costly programs. It is possible to link the reference tasks from some well-defined training domain to a set of national norms for the scales of that domain. Subsequent progress of groups of trainees on these scales then can be interpreted in relation to the national norms. An example is the NAEP's three adult literacy scales, which can be used to show progress in a variety of ways. Statistical methods for equating scales that include common anchor tasks have been developed and refined by large measurement organizations. Calibrated scales of mastery can provide policy makers in organizations with progress information and year-to-year trend information.

The NAEP adult literacy scales have the potential for becoming a part of mastery assessment systems associated with job training systems in a number of areas. By going to the expense of equating the three scales to the national scale, policy makers would be able to observe improvements in the general literacy skills of trainees on a national scale, and improvements in job-specific competence through other scales developed for that purpose.

Care must be taken to support the validity of inferences made from gain scores derived from a proficiency scale. (A gain score is calculated by subtracting the beginning score from the final score.) The use of gain scores often creates serious measurement problems. Mastery scales anchored with reference tasks offer a promising alternative to old-fashioned gain scores. The results of an evaluation need not be reported in terms of numerical values at all. Because the higher-level reference tasks refer to (and, it is hoped, predict) actual work competence, evidence that some number of learners are able to perform some difficult reference tasks well, and before instruction could only perform simple tasks, is enough. Experts will already have agreed that such tasks are very job-like and demanding, and such evidence is better than a decontextualized gain score.

Beyond Summative Evaluation

In the scenario, the concept of formative evaluation was demonstrated from the first intervention to the end. Evaluation was implemented as soon as possible into the ongoing classes and laboratory room, by introducing the mastery map and the initial small set of reference tasks. Measurement was used to improve the different components of the system incrementally.

A front-end analysis will usually show areas in the training that are not being covered by either class or laboratory instruction. New tasks can then be added to the instruction with confidence, to be taught initially by the existing instructors. New practice and feedback elements may be added after each pilot test until the reference tasks and their associated, instructionally oriented feedbacks come much closer to the goal of standing completely alone. When that goal is reached, the supervisor and fellow workers could guide a new worker through a series of reference tasks and lessons. The reference tasks would simultaneously enable the trainees to demonstrate new skills and progress and provide practice and instructional feedback. The supervisor and workers would still provide much of the context, the motivation, and the instructional management. They would do this as part of the ongoing work situation. Even after implementation is complete, a new technology- and measurement-intensive system should be used as a continuing performance- and productivity-improvement system.

If we plan from the first to use evaluation data to improve performance and productivity continually, no summative evaluation is needed anywhere in the process. Performance on reference tasks can be noted all along the way and, when anchored to a measurement scale, will generally provide more interesting and understandable information than score gains. Summative evaluation in training is, at best, a snapshot at a single point in time of one attempt to improve performance. The concepts of continuous measurement discussed in this chapter and embodied in the concept of mastery assessment systems make it possible to observe progress in groups or individuals at many points in time. We need no pre-set standard to judge our success, but can continue to achieve performance improvements through the use of continuous measurement. For example, if the management group in the scenario had set for themselves a target of 90 percent of the exemplary performance identified in the highest performers, this very target should later give way as improvements in technology and in the proficiency of the work force enable workers to set higher targets. Certainly such a continually improving system is an important response to the need for global economic competitiveness. The untapped potential for productivity improvement is there. Greenspan (1988) sees cause for hope:

> The potential for further gains in efficiency is immense. In part, it will require building on those adjustments that already have been put in place. But we also must take a broader view and begin to make adjustments that will allow more scope in the work place for individual initiative and enterprise. And it is time to consider alternative approaches to management and the organization of work (p. 10).

Conclusions

This chapter was designed to introduce training practitioners to some new concepts about how measurement science, revitalized by interactive technologies and by the disciplines of cognitive science, instructional science, applied artificial intelligence, and the studies of organizational change, can provide a new framework for assessing progress and can add new discipline to the development, implementation, and conduct of training. The context of this chapter is the economic challenge our nation currently faces—the demand for a much smarter, much more productive workforce that can deal with high levels of technology and automation flexibly and productively. In a nutshell, the message is that the nation's work groups can attain high levels of expertise by applying considerably deeper and more disciplined uses of science and technology and their handmaiden, measurement.

Deeper and more multidisciplinary methods of front-end analysis will be necessary. We need accurate models of the domains of knowledge to be traversed. We need human exemplars of high competence. We need maps of each domain to show the journey to the trainees and to track progress. We need reference tasks that allow performance to be measured at each stage of growth, from novice through intermediate to expert. We need cognitive and psychometric analyses to show the range of scales that span the domain. We need to understand the range of constructs, especially those that characterize expertise along the path, and to include deep analyses of constructs at the points of discontinuity where knowledge structures change in going from one level of expertise to another.

We need new methods of instructional development that are more iterative and that benefit from formative evaluation. Such methods of development must be tied closely to implementation plans and strategies. They must be designed to produce organizational change through an incremental strategy, rather than through the rapid and usually unsuccessful introduction of massive technological change in a resistant organization. An ongoing formative evaluation should produce incremental improvement, not only in the instructional materials, but in the implementation plans and tactics as well.

This chapter deliberately has projected a speculative and complex scenario, requiring a mixture of contributions from several different fields—training and training systems development, performance engineering, interactive systems applied to teaching and testing, knowledge engineering, applied cognitive science, instructional science, and educational measurement. The goal is to find a new mixture that will bring about massive improvements in performance and productivity through better use of technology, training, and implementation. Reaching this goal is not the task of measurement science alone, nor of training practice alone. It requires collaboration and cross-fertilization among disciplines in responding to the high challenge our nation faces. I hoped this chapter will generate a dialogue that will motivate greater contributions from each discipline and provide a basis for future combined efforts.

References

Bailey, R.W. (1982). *Human performance engineering: A guide for system designers.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bennett, R.E., Gong, B., Kershaw, R.C., Rock, D.A., Soloway, E., & Macalalad, A. (1988). *Agreement between expert system and human ratings of constructed responses to computer science problems* (Report No. RR-88-20). Princeton, NJ: Educational Testing Service.

Branson, R.J., & Grow, G. (1987). Instructional systems development. In R.N. Gagne' (Ed.), *Instructional technology: Foundations.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bruner, J.S. (1968). *Toward a theory of instruction.* New York: W.W. Norton.

Bunderson, C.V. (1987). The modeling of expertise. In *Proceedings of the Department of Defense/Educational Testing Service Conference on Job Performance Measurement Technologies.* Washington, DC: Office of the Assistant Secretary of Defense.

Bunderson, C.V., Inouye, D.I., & Olsen, J.B. (1989). The four generations of computerized educational measurement. In R. Linn (Ed.), *Educational measurement* (3d ed.). New York: Macmillan.

Carnegie Forum on Education and the Economy (1986). *A nation prepared: Teachers for the twenty-first century.* Washington, DC: Author.

Fine, S.A., & Wiley, W. (1971). *An introduction to functional job analysis: Methods for manpower analysis #4.* Washington, DC: W.E. Upjohn Institute for Employment Research.

Forehand, G.A., & Bunderson, C.V. (1987a, 1987b). *Basic concepts of mastery assessment systems.* Unpublished manuscripts, Educational Testing Service, Princeton, NJ.

Gilbert, F. (1978). *Human competence: Engineering worthy performance.* New York: McGraw-Hill.

Greenspan, A. (1988, January). *Remarks.* Paper presented at the Martin Luther King, Jr., Social Responsibility Seminar, Atlanta, GA.

Hart, A. (1986). *Knowledge acquisition for expert systems.* New York: McGraw-Hill.

Johnson, W.L., & Soloway, E. (1987). PROUST: An automatic debugger for Pascal programs. In G. Keasley (Ed.), *Artificial intelligence and instruction: Applications and methods.* Reading, MA: Addison-Wesley.

Kirsch. I.S. (1987, February). Measuring adult literacy. In *Toward defining literacy.* Symposium sponsored by the National Advisory Council on Adult Education, Literacy Research Center, University of Pennsylvania, Philadelphia, PA.

Kirsch, I.S., & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults—Final report* (NAEP Report No. 16-PL-01). Princeton, NJ: National Assessment of Educational Progress.

Lipson, J.I. (1987). Testing in the service of learning science: Asssessment systems that promote educational excellence and equality. In *Assessment in the service of learning: Proceedings of the 1987 ETS Invitational Conference.* Princeton, NJ: Educational Testing Service.

McCormick, E.J. (1979). *Job analysis: Methods and applications.* New York: Amacon, a division of American Management Association.

Naisbitt, J. (1984). *Megatrends.* New York: Warner Books.

Polanyi, M. (1962). *Personal knowledge: Toward a post-critical philosophy.* New York: Harper & Row.

Simon, H.A. (1981). *The sciences of the artificial.* Cambridge, MA: Massachusetts Institute of Technology Press.

Wainer, H., & Kiely, G.L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24,* 195–201.

Author notes

I am indebted to my colleagues at Educational Testing Service—Norman Frederiksen, Garlie Forehand, Irwin Kirsch, Myrtle Rice, Mike Rosenfeld, and Ben Shimberg—for providing valuable reviews and editorial contributions from the perspectives of measurement science, cognitive science, industrial psychology, and writing style. I would also like to thank Deane Gradous and Catherine Sleezer for their thorough editing. Any remaining mistakes or omissions are my own.

HRD and the Training Function

Is Training Really the Answer?

Randolph I. James
Consultant
James & Associates

In the chapter, "Measurement Science and Training," Bunderson suggests that a revitalized measurement science can help improve training. It is easy to agree with the basic inductive that he uses to draw the interest of the reader; "our country's ability to compete in world markets *is* being challenged"—and it *is* time to do something about it. A revitalized measurement science *can* help improve training, but it must be addressed within the context of a purposeful system. A purposeful system gains or maintains high performance from a front-end analysis that identifies and pinpoints the source of an improvement opportunity. This chapter describes how training and measurement science fit within this systems perspective.

Identifying Performance Improvement Methods

There are no magic solutions that cure all performance ills. Many methods are available and selecting the one(s) that will effectively address a performance improvement opportunity is not easy. Tom Gilbert (1978) suggests using six basic questions to help organize our identification efforts:

1. Do people have the *data* needed to perform well?
2. Do people have the *resources* needed to perform well?
3. Are there meaningful *incentives* and rewards?

4. Do people have the necessary skills and *knowledge*?
5. Do people have the *capacity* to perform well?
6. Do people have the appropriate *motives* and expectations?

Methods of improvement are identified by addressing each question in a systematic manner, in the order presented. This allows greater leverage in performance improvement efforts (i.e., the largest increase in performance for the smallest increase in cost). Figure 1 summarizes the areas to investigate when looking for methods of improvement.

Figure 1—Areas addressed by the six questions

DATA • feedback • expectations • guides, manuals • job theory • job aids	RESOURCES • work tools • work flow • environment • span of control • authority	INCENTIVES • monetary • nonmonetary
KNOWLEDGE • training • knowledge circles	CAPACITY • recruitment • selection • work schedules	MOTIVES • job redesign • career pathing • employee involvement

The methods that work in your organization will depend on available budget, support, and management policy. For example, there may be great potential to automate a task. However, if no budget is available to purchase the equipment then a different method would have to be pursued. Another method might require a decision to release more product information. If management will not support such a policy decision, then the method would not be appropriate.

Unfortunately, many managers consider training (aside from using more resources) as the only solution when it is just one of many. It is important to remember that training can be an effective method of improvement. However, it is only appropriate if the front-end analysis has pinpointed a deficiency in what a person should *know*.

The decision to pursue training is often determined by available budget and resources. However, training is only one component of the learning process. Job aids, procedures, and feedback are other low cost components that need to be considered. These components can also be used to complement a training effort.

Where Does Measurement Science Fit?

Bunderson's "four-generation framework" provides an excellent description of the applications of measurement science in the training environment. He points out that each generation does require a "technology that embodies increasing sophistication and power," however, the applications and resulting benefits are logical and well-presented.

Intelligent measurement and "mastery systems" in a computerized environment would certainly be great assets. But how practical is this, given the efficiencies of other nontraining improvement interventions? Intelligent measurement, no doubt, would be cost-beneficial where training is directed at a large target population (e.g., the military or educational community). However, most human resource professionals work with training groups that have "charge-back" budgets and that offer generic types of classes (e.g., basic supervision, interpersonal skills, etc.). Intelligent measurement in this type of training environment would be limited by development and implementation costs and the relative return-on-investment compared to nontraining options.

Bunderson suggests that performance engineering and measurement science can each benefit greatly from the findings and methods of the other. The primary lesson that performance engineering offers to measurement science practitioners is to apply their craft to the *total performance system.* This means looking beyond training and using measurement science to help do the following:

- develop user-friendly performance measurement and feedback systems;
- design more effective work process flows (e.g., through time series analysis);
- maximize the benefits of ergonomic/environmental factors that affect performance;
- develop cost-effective "reorder points" in supply departments;
- improve human resource recruitment and selection techniques;
- measure customer and employee satisfaction.

A more comprehensive list certainly can be generated. The point is that these and other nontraining areas represent additional applications for measurement science and offer benefits that will outweigh development and implementation costs.

Measurement science can make a significant contribution to the field of human resource development. However, it is a matter not only of improving *instructional* development but also of contributing to the emerging technology of *performance* development. The true charge of measurement science should be directed toward organizational performance. Training is just one of many subsystem components that need to be addressed.

References

Gilbert, T.F. (1978). *Human competence: Engineering worthy performance.* New York: McGraw-Hill.

The Role of Front-End Analysis

J. Diane Beezer-McClure
Rochester Branch Manager
Honeywell Protection Services Division
Honeywell, Inc.

The 1980s are years of dynamic competition. Off-shore manufacturing, mergers, acquisitions, and takeovers are happening in parallel with rapid technological evolution. Concurrently, the diversity of the onshore workforce is increasing significantly. American industry is responding to its challenges with a greater concern for increasing human productivity. As industry's attention is shifting in the direction of productivity, workers' attention is shifting toward desires for job satisfaction, challenging work, and compensation (Hartzell, 1988). Industrial survival, it seems, is dependent upon increasing productivity while at the same time providing workers with challenging positions that match their personal needs. To meet this goal, performance requirements need to be defined, and the performance itself needs to be managed and ultimately measured in terms of overall business objectives.

Against this backdrop, Bunderson's chapter, "Measurement Science and Training," speaks of significant gains to be made in human performance through the implementation of measurement science in training. He suggests that knowledge engineering tools and artificial intelligence should be developed and used in the application of measurement science to human resource training and that this will benefit both the organization and the individual worker.

Whereas Bunderson speaks of significant gains to be made in the HRD area, the reality of the present situation shows many human resource departments with computers still on the "wish lists." Bunderson is presenting the high end of a performance measurement continuum—the ideal world. Much of industry/business today is not yet equipped with the resources to utilize such technical expertise and innovation that would define all performance stages in a career-life cycle and provide real-time measurement of performance. However, progressive steps toward this goal need to begin now. HRD professionals can use Bunderson's ideas and begin to generalize the effects to performance management subsystems which include recruitment, selection, orientation, goal-setting, training/development, coaching, communication, appraisal, rewards, career development, and succession planning.

Performance management subsystems should not be islands of activity. The key to their success is through integration into the business. Front-end analysis of jobs is the beginning step toward subsystem/business integration. Dunnette (1976) outlines various methods for conducting front-end analyses of jobs (such as needs analysis, functional job analysis, task inventories), depending upon the type of information that is needed. Siegel (1987), responding to the need to manage performance on a broad basis, writes about the move toward integrating front-end analysis techniques to apply the results to all performance management subsystems.

Bunderson addresses several measurement science concepts in conjunction with front-end analysis of jobs—"mastery" beyond minimum competence; a "mastery map" of knowledge, skills, rules (formal and informal), job elements, and reference tasks to pinpoint developmental stages of performance; and summative evaluation. As mentioned earlier, someday the knowledge engineer will use artificial intelligence to track formal rules, informal rules, and facts to identify various performance stages. For now, HRD professionals can begin building this paradigm by expanding the process of front-end analysis of jobs so that it will

- identify the need for human expertise, and
- identify the specific knowledge, skills, and attitudes required to perform.

First, "mastery" beyond minimum competence should be pursued. Presently, workers are recruited, selected, and trained to be "typical" performers because expectations are projected from known "solid" workers. There is a need to go beyond defining jobs from incumbent and manager input to include (a) asking customers what service they desire to meet future needs, (b) searching for "experts" in the field, and (c) obtaining from original equipment manufacturers a definition of the most sophisticated use of machines—to name a few sources. In addition to searching for outlines of tasks performed and attributes needed to perform the tasks of a given job, "mastery" should be sought through (a) identification of product difference, (b) accuracy and speed of performance, (c) sequence of process procedures followed, and (d) supervision required.

Second, mastery maps should include knowledge, skills, rules (formal and informal), job elements, and tasks from various performance levels. During the front-end analysis, information should be collected from novices, advanced beginners, adepts, and experts to isolate differences among these performance levels. Once these differences have been isolated, performance management subsystems can be better tailored for the appropriate audience.

For example, a workforce should include workers in a variety of learning-curve stages. If the front-end analysis has previously identified the differences, other recruitment and selection tactics could be pursued as needed. Another example involves appraising workers. If workers at each performance level were given appropriate feedback for that level, acceptance of the feedback would be more likely to occur.

Finally, formative evaluation of jobs should be instituted to provide a continual data flow to keep the front-end analysis relevant in our change-based society. Content analysis of memos/reports, processes, and descriptive performance appraisals can show performance changes that have occurred in response to new business requirements. Another option would be to readminister task lists to incumbents and managers on a recurring basis to collect data on job changes.

Front-end analysis of jobs is the beginning step to managing performance. The use of technology can help to ease the recording, analysis, and use of job data. HRD professionals must continue to search for new definitions of "master" performers, seek job information from workers at various career-life-cycle stages, and pursue a constant flow of descriptive job data. The vision is to respond to the competitively dynamic environment, by optimizing measurement tools that positively affect worker productivity.

References

Dunnette, M. (1976). *Handbook of industrial and organizational psychology.* Chicago: Rand McNally College Publishing Co.

Hartzell, B. (1988). Job satisfaction is number one. *Personnel Journal, 67*(3), 6.

Sigel, G.B. (1987). Education and training for the job analyst. *Personnel Journal, 64*(7), 68–73.

The Training Function Within HRD

Gebreyesus Hamda
Senior Human Resource Development Analyst
Florida Department of Transportation

Human resource development (HRD) is a systematic effort to expand human potential and organizational effectiveness through a variety of interventions. Successful HRD efforts result in a well-managed workforce that can perform at higher levels of productivity. In this monograph, Bunderson ties deficiencies in the training of the U.S. workforce to the inability of this nation to compete in the world economy. He logically proposes that a trained workforce is likely to be more productive because training leads to the knowledge, skills, and attitudes necessary for better job performance.

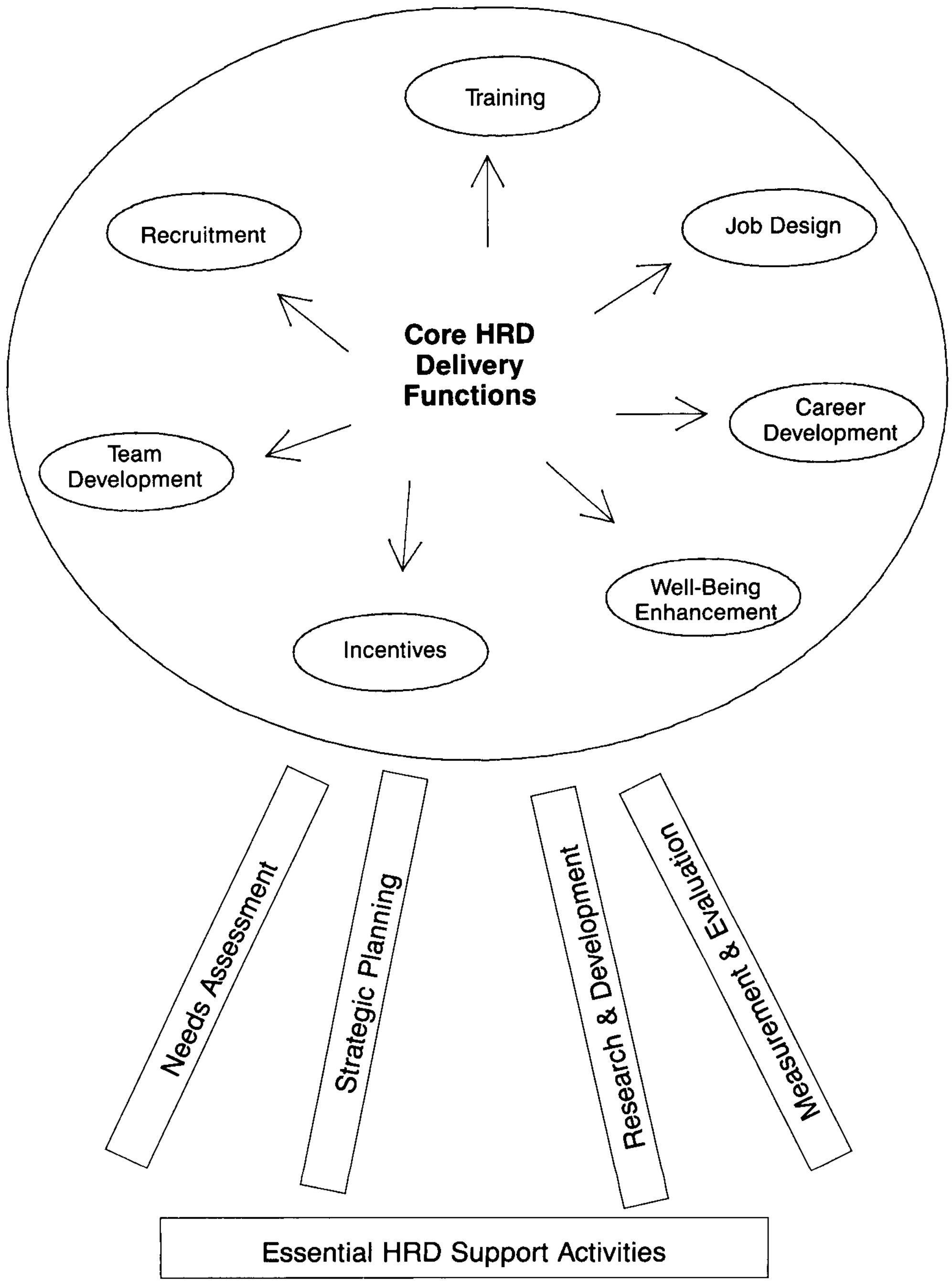
Training, however, is only one of the available HRD options. Some additional key interventions that affect job satisfaction, job performance, and productivity include recruitment, job design, career development, incentives, and team development. Most successful HRD departments also engage in support activities such as needs assessment, strategic planning, research and development, and measurement and evaluation.

Strategic planning helps forward-looking HRD managers prepare for the organization's and the employee's shared future. In Kaufman's (1987) terms, needs assessment provides the direction for useful problem resolution through identifying, documenting, and selecting appropriate (organizational and related human resource development) problems. Measurement and evaluation reveal the effectiveness of HRD programs. Research and development creates new knowledge and new technology that will help solve HRD problems. These activities are essential to assuring that core HRD functions are appropriate, effective, and of a high quality. It is important to note that although these activities are essential to effective programs, the activities themselves are not the core functions of the HRD program. Figure 1 shows the conceptual framework of the core delivery functions and the essential HRD support activities. This chapter explores the HRD intervention of training in terms of three features: domain, time range, and occupational focus.

Domain

Domain includes knowledge, skill, and attitude. Knowledge refers to intellectual awareness and articulation of concepts and facts. Examples of types of knowledge are knowing the capital city of the

Figure 1—A conceptual framework of human resource development



United States and knowing the differences between the HRD and engineering professions. Skill refers to the competence that enables a person to perform a given job function. Examples include driving a car safely, typing a manuscript correctly, or writing a computer program. Attitude refers to one's value system, such as one's likes, dislikes, and general dispositions toward an object, an event, a concept, or a person. For example, hard work is a value that most managers like to encourage in their employees; punctuality is a value equally well regarded by managers; the "just say no to drugs" campaign is a value statement that community leaders are promoting to the public.

The more domains of learning involved in training, the more meaningful the learning experience is for the trainee. Thus, if training efforts are to have lasting impact, attempts should be made to encompass all three domains. For example, a writing course could encompass all the domains by using the following objectives:

- knowing what the indicators of text readability are (knowledge domain);
- demonstrating the ability to organize a text (skill domain);
- choosing to use simple words familiar to the reader instead of big words (attitude domain).

Training Time

Both training and education can teach the same knowledge, skills, and attitudes, but the time requirements usually differ. Education generally takes place in a relatively longer time period than does training—that is, training is usually a short-term effort.

Because of this difference, organizations tend to use the training approach rather than the educational approach. Training is both a time saver and a money saver. Yet some organizations invest in educational programs because of their long-range payoffs. Such investments can make sense if complex skills, knowledge, and attitudes are required. Investment in educational programs is often required if a high level of expertise is demanded on the job.

Occupational Focus

Training has three possible occupational focuses: managerial, technical, and procedural.

Management training is designed primarily for managers to improve their competencies in such areas as planning, organizing, budgeting, directing, monitoring, and reporting. Typically, there are three broad levels of managers: first-line supervisors, middle-level managers, and top-level managers.

First-line supervisors are managers who function at an operational level of an organization. Their primary purpose is to ensure that employees produce a high-quality product or deliver a high-quality service on time. Managers at this level are generally new to management and, therefore, require training in fundamental managing skills.

Middle managers work between top managers and first-line supervisors. Typical responsibilities of middle managers include interpreting policy and procedures for first-line supervisors and managing other managers. Because a primary role of middle managers is communicating down to first-line supervisors and up to top managers, training programs for middle managers often focus on communication skills. Another focus is on program coordination skills because middle managers coordinate activities of different work units and of different managers and supervisors who report to them. In addition, because of their unique place in an organization, middle managers' overall management potential is often developed through educational programs to prepare them to fill top-level management positions.

Top-level managers are the senior executives who set the organization's policies. While managers at middle and first-line levels tend to deal with the organization's day-to-day internal activities, top level managers mostly deal with relations between their organization and the external environment and between the organization and the future reality. Thus, the training program for top managers generally focuses on policy development and strategic management issues.

Technical training is designed primarily for rank-and-file employees to develop competencies in their specific occupational areas. According to Birnbrauer (1985), obvious examples of employees who receive technical training include operators, apprentices, service and repair personnel, and clerks. Sometimes first-line supervisors and middle managers participate in technical training to hone their technical competencies. Areas of technical training in the Florida Department of Transportation include the following:

- microcomputer operations;
- graphic design;
- pavement markings;
- concrete structures inspection;
- signal system design;
- traffic noise analysis.

Procedural training is designed to allow both managers and rank-and-file employees to learn organization-specific procedures. Procedures for employee discipline, travel reimbursement, employee appraisal, and contract development and approval are examples of procedural training.

Managerial and technical competencies can be achieved either through educational programs or through training programs, depending on the level of complexity of the material and the time required to build these competencies. However, procedural competencies can almost always be attained through training. It is unusual for organizations to put their employees through educational programs to build competencies in company-specific administrative procedures.

In developing training programs, HRD professionals should always bear in mind the ranges of expertise (novice, advanced beginner, adept, and expert) identified by Bunderson. Regardless of the occupational focus, training is likely to serve participants at different ranges of expertise. Identifying participants' levels of expertise before and during training is crucial to designing, developing, and implementing a high quality training program.

Conclusion

The full range of HRD interventions has the potential to affect performance of an organization through both group and individual productivity and through employee satisfaction. Training is one key HRD function. It is important to note that the HRD functions are not the exclusive jobs of HRD professionals, but are, in fact, primarily the jobs of line management. The assumption is that HRD professionals advise and assist their organizations in these efforts. Measurement and evaluation, one of the essential HRD support activities, is an important element in advising and assisting organizations as to the conduct of their training.

References

Birnbauer, H. (1985). Technical training. In W.R. Tracey, *Human resources management and development handbook.* New York: American Management Association.

Kaufman, R. (1987). A needs assessment primer. *Training & Development Journal, 41*(10), 78–83.

SECTION 3

Measurement in Business

Human Resource Measurement in the Business System

Jeffrey D. Oberlin
Manager of Planning and Evaluation
Motorola, Inc.

Helping a business attain its goals is the mission of any successful training organization. Too many times training activities and business needs are not linked. This is even more obvious when one looks at the measurement systems many training departments use.

Whereas a business measures things like profitability at a corporate level, quality improvement at a divisional level, and productivity at an individual level, training often looks at something entirely different—the numbers of employees who have attended basic supervision courses, for example, or how many time management texts have been distributed to engineers. These training measures are clearly not in harmony with the business measures.

To improve business performance against goals, training professionals need to understand business measures at all levels and link training outcomes to business results. Training measurement systems should take into account the entire organization, not just its parts—the individuals, the work groups, the business units, or the products and services they produce.

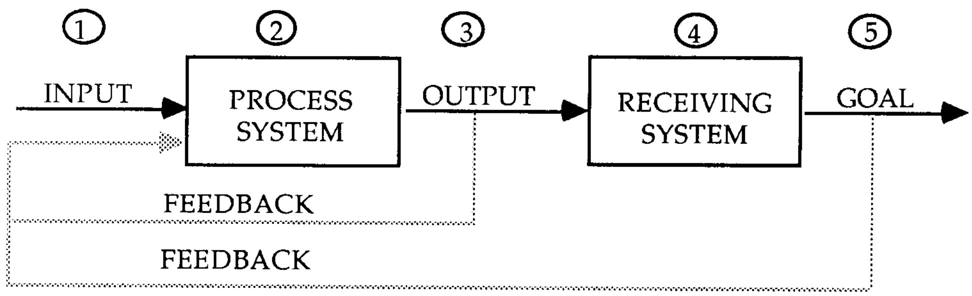
This chapter of the monograph reviews how training measures can more closely support the businesses they serve.

The Systems Approach

The systems approach to organizational analysis gives us a means to show the relationships among functions within a business (Brethower & Rummler, 1979; Gilbert, 1978; Rummler, 1977; Rummler & Odiorne, 1988). Based on the assumption that organizations act as systems, the training department must understand the system it serves and know where to place training measurements.

Figure 1 illustrates a traditional business systems model. Inputs are sent to be "processed," which creates an output that is "received," which leads to an organization's reaching its goals. This system incorporates a series of feedback loops that send data back to the processing activity. These loops are the mechanisms that inform a business about its movement toward its strategic goals.

Figure 1—Traditional business systems model



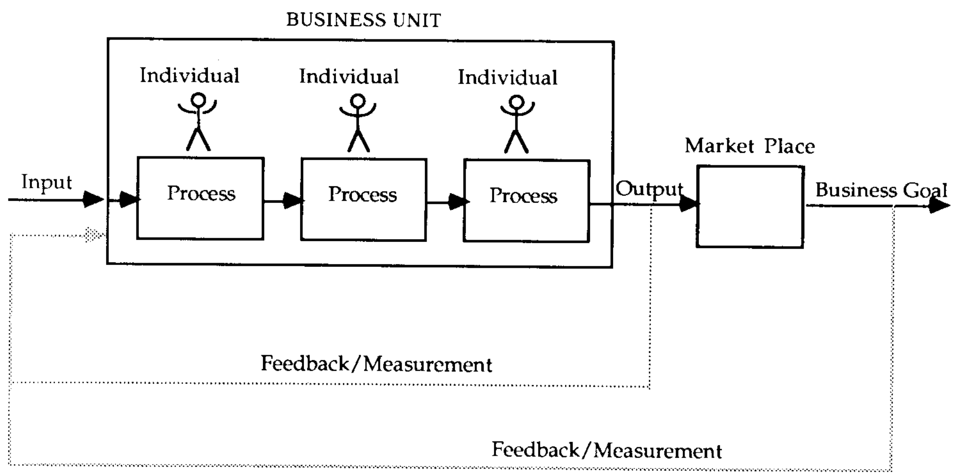
All businesses work as systems. They react to inputs such as vendor supplies and market demand. Throughout this system there are a number of points at which the business measures its performance. The numbers on the model in Figure 1 designate critical areas of business measures. Examples of these measures are as follows:

Critical Area	**Example**
1. Input	• The quality of materials supplied by vendors has to be sampled.
2. Process	• The number of insurance forms logged in and completed are checked daily.
3. Output	• The yields of integrated circuits are determined and checked against goals.
4. Receiving	• The time that it takes for a business to deliver its product or service to a customer is constantly measured.
5. Goal	• The movement toward the desired market share percentages is monitored by the business planners within the organization.

These measurements are set by the business. They are the critical areas of feedback for any organization. Because these measures are in place, training should align its feedback loops with those of the business. In this way, training will be able to affect the business' quality, cost, and time measures.

In the ideal system, the performance of individuals is linked directly to the organizational output. Figure 2 illustrates the linkage of individual performance to the overall performance of the business unit. Each individual is in charge of a particular process step that, when linked to other process steps, produces an output for the organization. The feedback to individuals on their performance must be based in part on the feedback received by the organization. If defective products are reaching the marketplace, the business and the individual have to know this. It is the only way they will be able to improve their performance.

Figure 2—Individual and business performance link



To improve individual and organizational performance, the training department must know what the business measures are, how these measures are affected by the business environment, and what impact training has on them.

Making Use of Current Business Measures

Businesses are continually changing the types of measures used on their products or services to assure success in the marketplace. A single measure or combination of measures informs the business about how well its product or service is performing. The following list identifies strategic business criteria and corresponding examples of product/service measures.

Strategic Criterion	Measurement
Cycle time	• Order entry processing • Equipment changeover time • First to penetrate a market
Customer satisfaction	• Repeat business • Number of customer complaints • Market share
Manufacturing excellence	• Benchmark against best in class • Field reliability
Quality improvement	• Reduction of defects and variation in products and services • Elimination of non value-added processes
Return on investment	• ROI versus other investment opportunities • Performance improvement from one year to the next

The growth and profit of a business are a direct result of how well its products or services do in the marketplace. Business management believes that the better their criteria and measurements are, the better their products or services will be and, therefore, the better their chances will be to beat the competition.

These criteria are not static. They are continually being changed by the internal and external environments. Typical environmental factors that affect the business measurement system are technological changes, resource supply, government legislation, and competition. The state of flux caused by the environmental impact on business is a major concern for the application of measurement science.

Training organizations that are able to adjust to rapidly changing business conditions keep the businesses they support competitive. As market demands change, the production or service process will also change. This process change, in turn, will modify the performance requirements of the individuals within each business unit. Any measurement science tool must be able to adjust to the evolving business scenario.

The following examples demonstrate the powerful effects environmental factors can have:

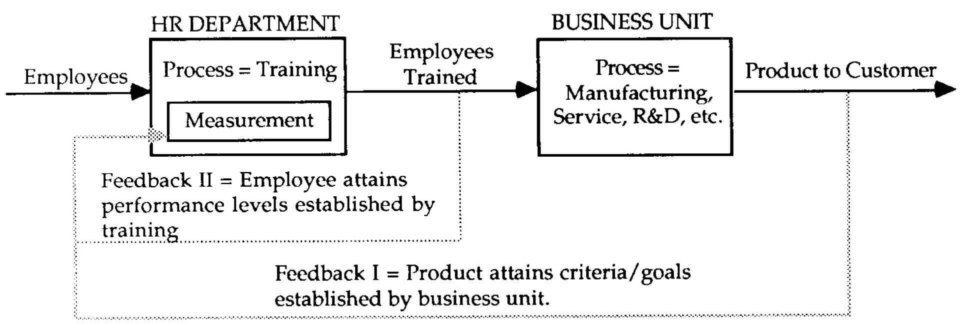
1. U.S. industries are now in a continuous struggle with European and Asian competitors. Globalization is a strategic imperative of many U.S.-based businesses. Company cultures are being transitioned to do battle with off-shore rivals. The changes required in

product design methods, distribution channels, and quality improvement are unprecedented. Strategically positioned training departments have been able to educate all levels of employees on the scope of this competition. In many cases these departments have had to teach senior managers the market share tactics of international businesses and, at the same time, train line workers on the cause-and-effect analysis used by the Japanese.

2. The current tax laws have made their presence known on a broad scope. Accounting and investment houses have been preparing and training for these changes over the past three years. To cover procedures on such items as capital gains reporting and depreciation schedules, businesses are monitoring rulings constantly. Training departments within these businesses have to modify their testing procedures every time a ruling is made. They also have to update all former students in the field. The training group unable to handle these requirements will soon find that the business' employees are making errors. In this industry, one error by one employee could cost millions of dollars in legal fees and lost customers.

The training model illustrated in Figure 3 shows how the output of the employees trained by the human resource department is received by the business unit. These employees in turn perform their duties within a process that delivers a product or service to the customer.

Figure 3—Training system model



The measurements in this system are on two levels. The first level is the measurement of the product or service produced by the business unit. If the business unit process is working correctly, then quality levels have been improved or maintained, cycle times have been reduced, and service support has been increased. The second level of measure is employee performance against training objectives that are closely aligned with the business unit's goals. Again, the process and the ideal setting will have the employees' skills training linked with the business unit process and its output.

Whatever measurement methods are used, the human resource department has to manage the linkage of employee assessment to the output of the organization. To do this, the training department should have three activities in place: planning and analysis, management involvement, and feedback.

Planning and analysis

Training must position itself strategically by analyzing the critical issues of the business unit (Kaufman, 1985). The department must determine the effects of these issues on key populations and develop training plans to address these needs. Individualized training cannot have an impact on the business if it is conducted in isolation. Plans have to be developed for all employee levels, functions, and departments. This is most critical in ensuring that training will improve the business unit's performance. The result is a training needs forecast directly linked to the critical business issues.

Management involvement

The management and line staff of the organization must participate actively in all training projects affecting their operations. By working closely with management, the human resource professional can determine the criteria that will be used to measure training efforts. For example, the field failure rate of a given product or the cycle time required to pay insurance claims could be used.

Feedback

The successful training manager must continually monitor a number of feedback sources to determine the effectiveness of training activities. More often than not, business unit management will assess the human resource department based on the perception of "doing the right things," rather than that of "doing things right." Therefore, if the human resources department's plans are driven by business issues, and management is part of all developments and interventions, then the chances are greater that training will help the business achieve its goals. The feedback measures of the ideal training system are as follows:

- **Results measurement: "Doing the right things."**
  When the business managers see the organization reach its goals, they *know* that training has served a major role in that attainment if training activities are linked with business unit strategies.
  The greatest results and payback for the human resource department are on the macro level. Has the sales force been able to satisfy the customer and generate more repeat business? Have order entry

control checks been eliminated as part of Quality Circles implementation? Has the factory reorganized to facilitate its "just-in-time" manufacturing strategy?

It is obvious that when the training process is part of the business process, the business unit's performance against its strategic criteria is the critical measure of training. This *is* "doing the right thing."

- **Results measurement: "Doing things right."**

Measurement of individuals needs to be in place so that training professionals know how programs are working, how objectives are being met, and how skills, knowledge, and performance feedback are obtained. Feedback/measurement systems need to be timely, understood, accurate, and accepted by the employee. If his/her skills have changed in accordance with established performance levels, then the training is considered successful for that individual. This is "doing things right."

All training organizations must have, as part of their ongoing accomplishments, activity that measures training performance on the individual level. Feedback data on this level is used by training management, instructors, developers, and employees. These data provide feedback on course design, instructor effectiveness, content acceptance, and individual performance improvement. Individualized measurements are necessary for a training department to work at optimal efficiency. Without these kinds of data, even though the department may be "doing the right things," its effectiveness will falter, and managers will question their training investments.

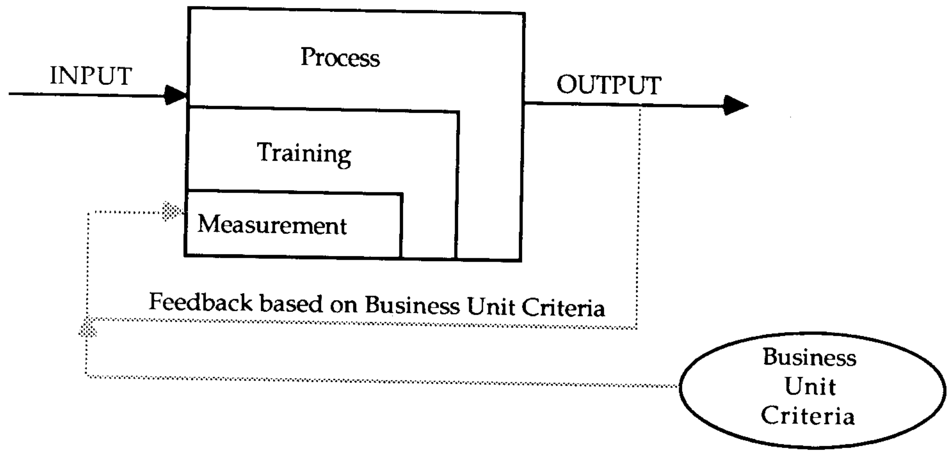
Continuous Measurement and Business Application Links: A Projection

Training has to be at the forefront of all business issues so it can correctly position its measurement needs. There are many sophisticated measurement tools currently available that can be used to determine training effectiveness, such as continuous measurement and behavioral analysis. But these systems should not be put in place until after training and business activities have been linked to align the business unit with job performance. The goal is to move simulation training to the job, where it can make adjustments for changes in business measures. This can be accomplished by taking the models presented earlier in this chapter and blending them with measurement science.

Figure 4 illustrates how measurement, training, and the business process can be integrated to improve business unit performance.

Intelligent computer-based measurement can integrate assessment tools, training, and the business process. Both the measurement and the training systems are embedded in the process and are somewhat invisible to the individual.

Figure 4—Embedded training and measurement systems



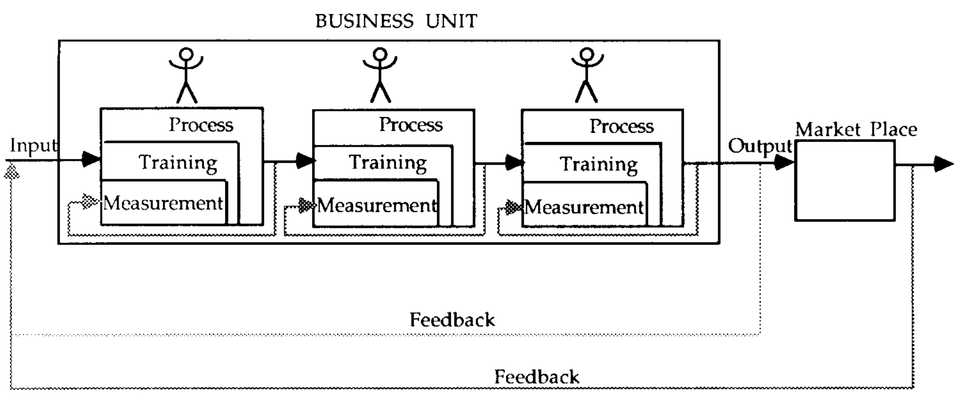
A convenient example of computer-based measurement application is in semi-automated manufacturing equipment (although order-processing and field service support could be applied as well). A business unit manufactures widgets, and the strategic measurement criterion used by the business is the quality specification of widgets going to the customer. Measurements are gathered on the employee's skills, knowledge, and abilities in working with the machinery and on the product's quality.

The interplay between the measurements and the training takes place within the machinery that produces the widget. The continuous measurement system in the machine will prompt/train the employee when either the skills, the product quality, or both are not in line with measurement standards.

Figure 5 depicts how the integration of training, measurement and business systems may interact. All measurement requirements have become a single system. The feedback needs of the individual, the business unit, and the human resource department are satisfied.

Will measurement science ever reach this level of sophistication within the business setting? That is difficult to predict. It cannot if human resource departments do not learn and prepare for future business issues, if analysis does not consider all environmental impacts, if the business and human resource department systems are not integrated, and if the human resource professionals do not have the capability to manage such systems. The deciding factor may ultimately come down to a cost/benefit analysis of such a system, and management will not pay if they do not know how the measurement process works. They need to understand the integration of training activities and how the business operates.

Figure 5—Individual and business performance link



They have to be involved in decisions of training development and associated measures. Once involved, managers will pay for elaborate measurements only where there is a decided competitive advantage.

It is obvious that the skills of airline pilots and nuclear engineers need continual measurement and training. High-fidelity simulators are commonly used for these jobs. The armed services also use the latest generation in measurement systems. It is easy to justify training expenditures when thousands of people need the same job skills. But industry today needs a flexible system that responds to an ever-changing environment. It also needs a system that can be applied to groups of less than one hundred employees. Few businesses have the inherent requirement of nuclear engineering training or the en masse opportunities of the U.S. Army. Therefore, common sense should prevail when the training organization proposes any type of measurement activity. The focus will be on what benefit the business will receive for the investment.

Conclusion

Before any training organization undertakes a move into new measurement technologies or into any type of measurement tools, the following strategic questions must be answered.

1. Will measurement be integrated into the full context of the business and training systems? The use of measurement data has to be strategically positioned with other training department accomplishments. The business unit has to place as much emphasis on measurement as it does on the planning, development, and implementation of training.

This will not occur without hard work and considerable time. Measurement is the feedback mechanism of the whole training process. Business management needs to know how to use these data to make informed decisions. When this happens, the importance of training results will increase.

2. Are the strategies of the human resource department aligned with the business unit's plans? All training activities, including measurement, should support the business goals. A sound training planning process will clearly define all the strategic criteria to be measured by business and training. The quality, cycle time, or profitability plans of the business have to be linked to the measurement plans of training.

3. Is the human resource department applying new technologies just to apply new technologies? Don't upgrade to a computerized training system before the measurement process is in place and perfected. This would be similar to implementing automated equipment before factory systems are stable. The process will not improve —mistakes will just be made faster.

4. Where is the human resource department positioned within the business hierarchy? No feedback data will ever be accepted by a business if it does not come from a credible source. This may be the most difficult growth segment for the training field in the next five years. When training measurements begin to be equated with financial measures in corporate board rooms, then the profession will know it has arrived.

References

Brethower, K., & Rummler, G. (1979). Evaluating training. *Training & Development Journal, 33*(5), 14–22.

Gilbert, T.F. (1978). *Human competence.* New York: McGraw-Hill.

Kaufman, R. (1985). Linking training to organizational impact. *Journal of Instructional Development, 8*(2), 23–27.

Rummler, G. (1977). You need performance, not just training. *Training, 14*(10), 5–53.

Rummler, G. & Odiorne, G. (1988). *Training and development: A guide for professionals.* Chicago: Commerce.

Education, Training, and Measurement Requirements for New Production Paradigms

C. W. Hoover Jr.
Executive Director
Interconnections and Power Technology Division
AT&T Bell Laboratories

This chapter focuses on the application of measurement science to new manufacturing processes currently being implemented in business and industry. The discussion covers the training needs associated with these processes, as well as the application of measurement science to improve the processes.

New Production Paradigms

About ten years ago, a new era in American manufacturing ushered in several important changes:

- complex computer programs for controlling existing manufacturing processes,
- elaborate means for storing and retrieving inventory, and
- robotics.

Although the application of these techniques has resulted in improvements in productivity, the net gain has not been sufficient to make American manufacturing fully competitive. In a renewed effort to gain a more competitive position, American business has been developing new paradigms and is now moving swiftly toward implementing production processes that are based on these paradigms. These processes are known by such names as quality, just-in-time/total quality control manufacturing, pull manufacturing, cycle time control, and flexible manufacturing. Often these processes are combined in a "product realization process" within a company.

The scope of these new production paradigms goes well beyond manufacturing. Indeed, these techniques apply to aspects of work ranging from order entry through delivery, and they are particularly well suited to the development and manufacture of products that have a wide range of features and functions. They are also appropriate to the short development cycle times that are required in today's competitive environment.

T.A. Kochan (1988) characterized the nature of these processes and the changes in employee participation as follows:

> The more meaningful improvements in economic performance are being achieved in settings where quality circles or related forms of employee participation are integrated in much broader changes in manufacturing processes and strategies that emphasize a commitment to quality, flexibility, and continued improvements in productivity. Where combined with these new manufacturing systems, employee participation and industrial relations reforms have also been broadened to encompass work organization changes, expanded training, more decentralized decision making, and, in general, greater flexibility in the use of human resources.

Implementation of these new processes radically changes the measurements needed in design and manufacturing. In design, quality and manufacturability are measured. For example, specific techniques for rating the manufacturability of a new product design are essential elements of the product development process.

In production, the processes are measured—not the workers. For example, workers are charged with taking measurements, fixing problems whenever possible, and calling for help when needed. More workers mean more minds, eyes, and ears available to detect problems. Thus, detection and control are continuous rather than sampling processes. An important result of this change is quicker response to problems. Production lines are physically arranged (such as in "U-shaped" horseshoe patterns) so that adjacent workers can communicate easily. In this setting, local load balancing, fault detection, and correction naturally become cooperative efforts. In-process checks, which in the past were staunchly resisted by manufacturing engineers, are easy to implement and afford fine-scale process control. With this new mode of operation, control of the line is vested in the operators. For example, under the doctrine of pull manufacturing, the workers are expected to hit the red button to stop the line whenever faults occur.

Table 1 contrasts job assignments under traditional production processes with those under new processes.

Table 1—Comparison of job assignments

Function	Traditional Production	New Paradigms
Quality	Quality control, quality assurance personnel	Production workers
Machine maintenance	Maintenance force	Production workers
In-process checks	Manufacturing, engineering (if done at all)	Production workers
Fault detection	Quality control, quality assurance personnel	Production workers
Work-load balancing (fine scale)	Manufacturing, engineering	Production workers
Production data	Manufacturing, engineering	Production workers

With the new production paradigms, workers have a much greater opportunity to use their innate skills and intelligence to perform more interesting jobs. Cross-training, another element of this new mode of operation, both contributes to efficient production and enriches jobs by introducing variety. At a 1986 IBM conference, C.R. Frost described to manufacturing executives some of the advantages of cross-training (Frost, McIntyre & Nauta, 1987; Mitchell, 1986). Frost helped make his company fully competitive by implementing the processes, education, automation, and cross-training called for in the new paradigms. His total workforce remained constant, but many of the traditional jobs were replaced by better, broader ones. His company now operates 24 hours a day, 361 days per year. He provides for vacations, sick leave, education leave, and other absences by drawing on the ranks of his cross-trained workers. He reported that everyone finds this an efficient, exciting, and rewarding way to operate.

New Product Realization Processes

The investment required to implement the new product realization processes is principally in education and training. In sharp contrast to earlier techniques, these processes are not capital intensive, and successful implementation often leads to reduced capital requirements. Three types of education and training are required for the modern product realization process.

Cultural change and education. Increased cultural awareness in the organization and buy-in by all employees through education is necessary for the radically new processes to take place.

Team-building education and training. The new approaches work only if there is shared responsibility for the success of product offerings. Marketing, design, and manufacturing must all contribute. Appropriate education and training will help break down barriers between these functions.

Functional education and training. Teams that carry out the processes must be educated in specifics, from design for manufacturability to principles of statistical quality control.

Of the three types of education and training, functional education and training best fits Bunderson's measurement science theory, as he describes it in this monograph. His vision of education and training— resulting in increased productivity—has the potential to be realized in the manufacturing and service sectors. Many thousands of employees in both sectors need quality and productivity information that is highly specific, standardized, generalizable and critical to the economic survival of American business and industry. The large number of workers who need to learn the new methods may well make Bunderson's high-cost training and measurement proposal feasible.

What measurements of human performance will be needed? A complete answer to this question will require careful study by experts. However, it seems clear that at least the following types of measurements will be needed:

- for process developers and product designers—measurements of the ability to communicate and work with all the people involved in product realization.
- for the product realization teams—measurements of the ability to cooperate, adapt, and be innovative in problem solving.

Are U.S. workers willing to adopt the new modes of operation? Kochan recently noted that survey data consistently report that more than 80 percent of the workforce express an interest in gaining more influence over job-related decisions that allow them to use their skills more fully. Kochan also noted that there is little evidence that employees resist the introduction of new technologies or that there is a shortage of employees able and willing to be trained in the use of new manufacturing technologies.

In summary, the new production paradigms and product realization processes radically change the nature of jobs and, therefore, education, training, and measurement requirements. Education and training are essential to bring about the required cultural change and to provide the specific skills needed to make these processes work. In this new mode of operation, the product quality and other characteristics of the processes, rather than the performance of individual workers, are measured. In addition, new measurements of human performance, such as those proposed by Bunderson will be needed as guides to selecting and training the teams.

References

Frost, C.R., McIntyre, R.H., & Nauta, R.W. (1987, March). Developing a manufacturing strategy for the smaller company. *Production*, pp. 2–7.

Kochan, T.A. (1988, April 14). Adaptability of the U.S. industrial relations system. *Science*, pp. 240–287.

Mitchell, R. (1986, January 27). Automation is not just for the big guys anymore. *Business Week*, p. 52.

Integrating Measurement and Training Systems into the Organization

Michael T. Wood
Manager, Training & Technical Assistance
Industrial Technology Institute

Joseph T. Martelli
Research Scientist
Industrial Technology Institute

Nancy R. Mosier
Manager in Education Consulting
Arthur Andersen & Co.

This chapter discusses some issues and opportunities for bringing measurement systems, such as the one proposed by Bunderson, into the practical world of industrial and business training. The first section describes the need to plan for implementation, with emphases on models of organizational change, goals for a measurement system, and justification and support. The next section discusses key points for implementing critical components of the measurement model. Additional sections are devoted to actors and their roles in a measurement system and then to promising opportunities for implementing measurement systems. The chapter closes with discussion of the need to monitor and maintain the measurement system.

A Model of Organizational Change

An organization that applies the concepts of measurement science to training is changing the way it does business. Employees will be trained and performance monitored in more accountable ways. Management will receive and interpret new reports. Measures of performance and productivity will be more easily attainable.

To achieve management buy-in and employee support, the system must be implemented slowly. Management must show success in one or two areas and build on that success. In addition, the system must be seen as nonthreatening and as an opportunity to improve both individual and organizational performance. Implementing such an approach requires a change strategy that meets the needs of the people involved. This section reviews a model for planning and managing change that has been used successfully in many situations.

The Stages and Strategies of Implementation

During the successful implementation of change, employees go through five stages (Dormant, 1986): awareness, self-concern, mental tryout, hands-on trial, and adoption and integration. If employees have positive experiences and impressions during these stages, they will carry positive messages back to fellow employees, and implementation will be much easier.

Awareness

In the first stage, employees have little information about the system being implemented and tend to be passive in their approach to it. Managers or others involved in implementing the change must provide positive information that is general and brief.

Self-concern

As employees become more aware of the system, they will start to wonder how their jobs, the work organization, rewards, and incentives will change. They will begin asking questions such as, "How will management measure productivity?" and "Who will see my performance data?" During the self-concern stage, managers should encourage employees to express concerns and should be prepared to give direct, reliable responses and to be sympathetic and reassuring. Management may also prepare short presentations or seminars to show sample training activities, explain how data are collected, provide sample individual and aggregate performance reports, explain who will have access to the information, and answer questions.

Mental tryout

Once personal concerns have been addressed, employees will start to imagine how work tasks will be done after the implementation. This is the mental tryout stage. Management can help by providing videotapes of successful work groups, testimonials, and presentations. The mastery maps and other materials received during presentations may enhance mental tryout.

Hands-on trial

This stage is the time for management to provide actual instruction on the use of the new system. Instruction should include explanations, demonstrations, site visits, and opportunities for practice. This stage is an opportunity for managers to show how they adjust instruction, measure progress, provide feedback, and monitor performance over time. For employees, the hands-on trial is an opportunity for them to use new equipment, enter and retrieve data, and see new reports. This stage is the first opportunity for employees to see how the mastery assessment system really works.

Adoption and integration

Thorough instruction and realistic opportunities for practice facilitate the final stage. Here, the employees begin to use the new system and incorporate it into their daily routines. Management should provide ongoing support to users and should monitor successes, problems, and overall performance data.

Positive experiences in the classroom will facilitate adoption. Ongoing monitoring of employee acceptance of the measurement assessment system is essential. Because most work groups generate and review performance reports regularly, supervisors can and should ask employees what they like and dislike about the system. Likewise, managers should ask supervisors what is working well and what is not.

Measurement in the Context of Phased Change

There are parallels between the stages of implementing a change in a training system and implementing a change in technology. Bunderson's model for training follows the framework of instructional systems development (ISD), which consists of five phases:

- front-end analysis;
- design;
- development;
- implementation; and
- continuing evaluation.

Similarly, the process of technology change can be conceptualized as a phased series of activities (Wood, 1988):

- diagnosing the technology and the organization needs;
- planning the change;
- justifying the plan;
- selecting specific technology components;
- implementing the technology;
- gathering feedback and evaluating the implementation.

Both models involve measurements of needs at the start and outcomes at the end. Measurement variables and instruments can be tailored to gather information about what training is needed and how it has worked. The front-end analysis is an opportunity for management to gather extensive data about entry-level skills, job tasks, and expert behavior. This phase is also an opportunity to introduce the goals of the system and to analyze the development of both novices and experts. Involving employees in the front-end analysis is also a way to increase their acceptance early in the process.

Research conducted at the Industrial Technology Institute has resulted in standardized (and "software-ized") instruction for analyzing technology needs. Currently, the Institute is appending instrumentation for analyzing training needs. This type of conjoint measurement allows training needs to be identified in the context of the tools with which people will be working.

Evaluation requires continuous feedback on the implementation, or, to use Bunderson's term, continuous measurement. Both individual progress toward mastery and organizational performance should be measured. This can be tricky. The standardized measures available for assessing effectiveness, knowledge, and problem-solving skills are apt to be more limited in a setting that emphasizes both individually selected learning and organizationally tailored training. Feedback measures may more appropriately involve scaled work observations or task simulations that the individual performs. Although evaluation of training is uniformly recognized as important, measurement of training tends to lag behind measurement of needs assessment.

The intermediate phases, from planning and design through implementation, require new attention to measurement tools, if measurement is to be continuously embedded in the training and change system. If an effective needs analysis/task analysis has been done, then criteria can be set for selecting technologies and training approaches. Such measurement would address both training's organizational fit (sociotechnical context) and the individual fit (knowledge and skills requirements). However, for measuring the ongoing process of training qua the mastery model, new, efficient, and unobtrusive measurement schemes will need to be devised. This will be especially important for calibrating reference tasks and designing performance indicators for individuals following potentially different paths to mastery.

Figure 1 shows the the needs and opportunities for measurement during phased change.

Figure 1—Needs and opportunities for measurement application

Phase of Technology Change	Training Process	Measurement
Diagnosis	Front-end analysis	Needs assessment and task analysis
Planning	Design	Skills inventories and training needs
Justification/selection	Development	Cost-benefit and organization fit
Implementation	Implementation	Efficiency and performance
Evaluation	Evaluation	Outcomes and process

Goals for a Measurement System

A measurement science intervention can serve a number of purposes within an organization. Such a system can serve as an alternative or addition to a traditional training and development function. A measurement science approach to learning will never replace the entire scope of training and development, but it affords the opportunity to show that the training and development function can play a vital role in an organization.

A measurement science system is directly related to, and can contribute significantly to, an organization's performance and effectiveness. The component criteria of organizational performance become the reference tasks of the measurement science system, which is more like a performance system than a training system. Although some training and development departments are performance-oriented, many are not. Many operate as internal schools offering a host of generic courses (supervisory training, time management, project management, listening skills, etc.). Most are only vaguely tied to any predetermined or desired mode of organizational performance. In fact, D.L. Georgenson (1982) reported that of the $100 billion spent annually on training and development, less than 10 percent results in skill transfer to the job.

A study of one organization by J.T. Martelli (1987) found that the traditional approach to supervisory training was insignificantly related to job performance and, under certain conditions, was inversely related to performance. Because many organizations offer training as a reward or incentive, as opposed to a method for compensating for a performance deficiency, it really does not matter that the courseware isn't referenced. Such training serves as psychological M&Ms more than as an integral part of a performance system.

Unlike unstructured training interventions, a measurement science approach is strongly and positively correlated with job performance. Consequently, training efficiency and effectiveness are improved. Classrooms can be reduced or eliminated, and scheduling problems associated with training disappear.

A measurement science strategy is effective not because it's new or computerized, but because the development process is analytical, and good analysis produces valid training. Analysis of a measurement system is similar to the kind of analytical front-end work that goes into preparing any efficient and effective training intervention. In a traditional ISD model, this process involves a variety of structured analyses (task listing, task analysis, process analysis, troubleshooting analysis, etc.). In a measurement science intervention, the equivalent of these processes is embedded in reference tasking. What it all boils down to is an accurate and sufficiently detailed approach to symbolically representing knowledge and skill, perhaps to a higher level of expert knowledge than the typical trainer is used to documenting. As in any structured training intervention, the front-end work may appear costly and labor intensive. It generally proves to be cost-effective, however, because the needs analysis has already addressed training feasibility.

Structured interventions are not always successful. Failure is usually due to one of two problems—a faulty needs analysis or a faulty training analysis (typically the former, and not so much faulty as inadequate). The distinction is important. A faulty needs analysis generally results in macro-level problems (such as a training program for a problem that is not a skill or knowledge deficiency, the wrong program, or the wrong delivery system), whereas a faulty training analysis (task analysis, task listing, process analysis, etc.) will result in invalid content being delivered. Training analysis, being the more well-defined of the two processes, is usually not the problem. It is far easier to be misled in the needs analysis process and, consequently, to recommend and pursue an inappropriate solution.

Justification

Implementing a measurement system can be costly. The justification must be based on business measures that are commonly used by upper management. This section reviews six approaches to cost justification (Wang Laboratories, 1985) and illustrates how they apply to the implementation of mastery assessment systems. The approaches are: cost displacement, cost avoidance, strategic opportunity, financial analysis, value added, and technological imperative. This section also discusses flexibility in the unit of analysis as a justification of mastery assessment systems.

Cost displacement

Cost displacement refers to the savings that result from reducing or reallocating existing costs. It usually refers to money already being spent by a company and is often associated with the automation of repetitive work. In Bunderson's hypothetical case, substituting the mail-order software for manual procedures is an example of cost displacement.

Cost avoidance

Cost avoidance means eliminating future costs. In Bunderson's hypothetical case, the front-end work and continuing evaluation yielded speed and error rates that became standards for trainees and were incorporated into the mastery assessment system. Employees were trained more quickly and accurately and were able to return to the job environment with high-level skills and abilities. They were able to save the costs associated with slow work patterns and high error rates.

Strategic opportunity

This is an entrepreneurial approach to cost justification that creates business opportunities through the strategic use of products, services, or systems. Imagine, for example, a company that sells a complex software package. If that training package were supported by training that was based on expert behavior and included ongoing evaluation and feedback, the company might use the package to differentiate itself from the competition. Reduced training time and improved accuracy and productivity reporting might be particularly appealing to organizations that do not have the time or resources to develop training and that are interested in ongoing reporting systems.

Financial analysis

Corporate financial officers analyze the initial and ongoing costs and the immediate and long-term returns of potential investments to determine which will yield the greatest returns over time. The most common financial analysis techniques include payback period, average rate of return, net present value, cost-benefit ratio, and internal rate of return (Gurnani, 1984). Although often used to analyze capital expenditures, these techniques are not often used to make decisions about training investments (Mosier, 1986).

In Bunderson's hypothetical example, management requested an economic analysis of the payoff from a higher level of expertise. Instead of using performance potential ratios, they could have used financial analysis to compare the costs of implementing the measurement assessment system and the costs of letting the workers continue to make errors, produce work needing revisions, and work slowly. The resulting figures are typical measures of corporate performance and are familiar to financial officers and comptrollers.

Value added

The value-added approach defines benefits in terms of qualities rather than quantities. However, many value-added benefits, such as time savings and reduced turnover, can be expressed in monetary terms.

Some value-added benefits for training, and for mastery assessment systems, include time savings, better service, better decision making, reduced turnover and absenteeism, and job enrichment. In Bunderson's hypothetical case, time savings and better service were achieved by reducing errors, revisions, and slowness. Better decision making occurred because management received aggregate data to monitor productivity and work groups received individual data to help raise individual performance levels.

Technological imperative

Technological imperative is an approach used by those who are late adopting a new technology. It refers to a sense of urgency that if the technology is not adopted quickly, the company stands to lose market shares and competitiveness. The technological imperative also may occur when some work groups have adopted a particular process and others have not. If a few work groups implement a measurement assessment system, others may wish to follow suit, either because they fear losing productivity ratings or visibility, or because adoption is easier when they can see the benefits and successes in another group.

Unit of analysis

The approaches outlined can be used to justify measurement science and mastery assessment systems in terms upper management or a comptroller would understand and appreciate (i.e., the relationship to the bottom line). Another justification is the ability to report data at various levels. A well-designed mastery assessment system must collect data on individual trainees or users, but it should also provide aggregated reports that are appropriate for various levels of the organization.

For example, individual trainees should be allowed to see data on their own progress and performance. Work group supervisors and training administrators may also wish to see data on individuals, in order to provide better instruction and mentoring. Supervisors may use group-level data to analyze and improve group performance. Data may be aggregated further for upper levels of management who need to see only department or division performance reports.

Bunderson's hypothetical case shows how data on individual performance were made available to individual performers, to their quality circles, and to their supervisors. Those data were used to raise performance levels, reward productivity, and improve the system. Management used aggregate data on work group performance to monitor the system and establish performance standards.

Implementing the Measurement Model

Implementing the measurement model requires considerable forethought, resources, and organizational support. Although Bunderson's model comprises several important concepts and practices, three are clearly central, new, and difficult to put into operation. The three key elements—reference tasks, continuous measurement, and formative evaluation—require implementation support in terms of systems, equipment, and people.

Reference tasks and calibration

Effective mastery training requires identifying reference tasks and calibrating (scaling) performance levels for those tasks. The key support requirements for gathering and using that information involve a system for internal organizational analysis and networking.

The internal analysis system refers to ongoing monitoring of required work tasks and levels of skill in the workforce. Updated job descriptions and task analyses will be required to know what tasks must be calibrated for measurement. These tasks may be referenced either to jobs or to technology systems people operate and maintain. Task referencing would accommodate changing technology (as new systems or components are added) and changing jobs or roles within the factory or office. The monitoring of the workforce will be important to determining, at any given time, workers' knowledge, skills, and level of mastery. In turn, monitoring will require dedicated effort by the human resources staff or plant management.

Networking is likely to be important for deriving useable scale values for task mastery. Networking means collecting information from a variety of organizations where similar tasks are performed. Practically, this may be difficult to do, given the proprietary and competitive nature of private organizations. However, it may be possible to network informally among human resources or training collaborators, to develop a centrally monitored and widely accessible database of calibrated reference tasks (by job type, industry, production system), or to provide task data through a professional organization.

Why is networking such an important support requirement? Many jobs have a relatively small set of task requirements and underlying knowledge and skill foundations (such as data-entry or lathe operation). But for more complex jobs, such as programming a manufacturing process, monitoring a control room in a continuous-process production facility, or designing and engineering a new product, there are many more task and knowledge requirements. Work-specific tasks are more difficult to calibrate and generalize than, for instance, the more generic and basic NAEP adult literacy skills cited by Bunderson. Still, basic skills are fundamental to technology-driven work tasks, and the adult literacy model might be extended, at least, to an analysis of "technology literacy."

Continuous measurement

Continuously measuring individual progress toward mastery is another central element of Bunderson's approach. There are two crucial implementation requirements that will allow this to happen. First, the system needs a set of tiered, repetitive, applicable measurements of training effectiveness in terms of individual knowledge and skills. Those measurements must operationalize the underlying reference tasks and knowledge. The indicators could be paper-and-pencil or computerized tests, task simulations, or post-training observations of work performance. The indicators must be easy to use and diagnostic of future remedial or advanced training.

Second, implementation of continuous measurement requires an organizational system for using the measurement results regularly. Most formal performance appraisals occur only once or twice a year, but whenever measurement occurs, individual results must be communicated. If the measurement information is to be used for upgrading the training program, then a system for keeping records, analyzing data, and auditing the training will be needed. Moreover, the concept of continuous measurement can be construed in some work environments as a real or implied threat to the employee. Particularly in unionized settings, where there is concern about "surveillance," direct or unobtrusive measurement must be justified, constructive, and accepted.

A corollary requirement for maintaining a continuous measurement system is the equipment for measuring and giving feedback. Particularly if the training is to be programmed or self-paced, some computer-based training is a sensible delivery approach. Computer-based training could be used for structured or self-administered measurement. Training modules could be followed with tests of knowledge and immediate and private feedback to the individual. However, such a system requires extensive hardware and software. Some large, training-progressive companies have personal computers at every desk and ongoing training opportunities for employees. Other companies have established "learning centers" with training equipment, computers, and other learning and measurement aids in a centralized location away from the work station. Smaller and less affluent firms may be less willing to commit the capital or space required of those approaches. Yet if the decision is made to pursue training measurement, the alternative would involve costly, labor-intensive training delivery, assessment, and feedback. The costs and benefits must be weighed carefully.

Formative evaluation

A third basic ingredient of Bunderson's measurement model is formative evaluation, which means that evaluation of the training is ongoing and occurs at various stages of the training program. The focus is oriented more toward internal processes than toward end results, and measurements are used to fine-tune the evolving implementation. The evaluation forms itself, rather than being a slice-in-time assessment of impact.

The concept of formative evaluation, or formative validation, has been around for several years (Bowering, Fortune & Wood, 1985), and its value has been argued elegantly. However, applications to training programs are rare. The concept has been applied only in a few organizations where task analysis, training, and measurement are extensive, long-term activities. Formative evaluation is potentially very valuable, not only where processes may change, but also where skills requirements may change with the adoption of new technologies. Whether advanced technology "upskills," "deskills," or adds qualitatively new skills requirements (Jacobs, 1987), the evaluation should change with the changing expectations for the workforce.

Implementing formative evaluation requires measures of task efficiency as well as mastery attainment. Interim feedback on mastery will help identify further or refresher training needs.

Bunderson argues that formative evaluation can replace summative, or end-result, evaluation. In his terms, it is as useful to know that a certain number of people can perform a task with mastery as it is to know individuals' gains in knowledge or skill. His point is valid in the case of traditional before-and-after comparisons of performance. However, both to sell the measurement approach to management and to assess the overall training and development effort, some form of summative, bottom-line evaluation is still worthwhile and important.

Thus, a complete evaluation package should contain

- formative evaluation of individual learners via measures of mastery attainment;
- formative evaluation of the training process via measures, observations, and indicators of time and cost efficiency of the training; and
- summative evaluation of the impact of training and mastery attainment on the organization's productivity and costs.

Actors and Roles

Effectively implementing a training measurement system requires attention to the people who will be doing the implementing. A measurement system, in effect, gives new responsibilities to trainers, trainees, and company management. Although typical training programs involve some form of measurement, the mastery assessment approach involves new considerations for a variety of functions: task analysis, test construction or acquisition, measurement scaling, measurement computerization, system implementation and administration, data analysis and interpretation, communication and feedback, and coaching and mentoring.

Multiple expertise

Implementation requires many kinds of expertise within the organization. In particular, implementation requires advanced skills in measurement and integrating measurement with both training and the work environment. This expertise must come either from employees who are highly educated or trained in the educational or organizational measurement sciences or from consultants who can be brought into the organization to design and test the system and train the trainers to use it.

The players involved include people who train (or develop the training materials), subject-matter experts, trainees, people who use the training results for personnel decision making, and trainees' supervisors. All players in the training measurement system have responsibilities for applying their expertise to make the system work. The trainers and assessors must deliver and measure the training consistently and effectively. The trainees can use the skills they bring to the job to understand the training and apply it on the job. Management staff should reinforce mastery with job-related rewards that allow people to use their heightened skills. Supervisors need to be familiar with the training, tasks, and mastery levels of their work groups to facilitate the transfer of training and to coach their staff through successive levels of training to mastery.

Development and implementation teams

Because of the roles of multiple players involved, effective implementation requires a team approach within the organization. In addition, the concept of a training measurement system will be new to most organizations, and it must be understood and accepted. Such a system cannot be singularly conceived and implanted in an organization.

Educational team exercises or simulations can help prepare companies to handle implementation issues. For example, product and process designers and engineers have worked in teams to design products for manufacturability (Wood & Stoll, 1988). Managers, supervisors, and engineers have shared experiences playing the Advanced Manufacturing Technology Implementation Game (Vasilash, 1987), which uses roleplays to bring about technological change in a production company. Similarly, a team of knowledgeable and affected parties in a firm could design mock or real applications of measurement science to meet training needs. Such a team approach would, of course, involve substantial commitments of time and effort.

Rewards and incentives

Incentives are necessary to make a measurement system work. It might be argued that most of the incentive value should be intrinsic —that is, the training and resulting mastery allows trainers and trainees to develop professionally, use their talents, derive more satisfaction directly from doing their jobs, and be upwardly mobile toward jobs requiring higher levels of knowledge and skill. However, extrinsic incentives also are needed to motivate both learners and trainers to participate in and gain maximum benefit from training and measurement.

A mastery assessment system increases the load on the trainer, who has to design for, and train to, mastery and perhaps conduct the measurement. The results of the trainer's efforts are also more visible when measurement is continuous than in traditional systems. Trainers could be motivated with recognition for their expertise, additional pay, advance notification of schedules, simplified procedures for administering measurement, and benefits to their units (Shikiar, Wood & Saari, 1985).

Extrinsic incentives for learners need not be costly. For example, in a military setting involving technical training in electronics, mechanics, and similar areas, social incentives were tied to training success, defined by completion time and test scores (Wood, Klimoski & Del Gaizo, 1974). The values of an array of potential incentives were first determined by scaling their importance to the training audience, and the more valuable incentives were then used to reward successful trainees. The incentives included various forms of official recognition, assignments involving peer support and leadership, and time off (passes, reduced work assignments, etc.).

In implementing a training measurement system, a combination of valuable and manipulable incentives should be considered. For both trainers and learners, the question of "what's in it for me" will be answered with evidence of measurement benefits to individual and organizational performance and development.

Opportunities for Implementation

To stimulate the expanded application of measurement science to training, it may be helpful to suggest areas where a measurement system is most appropriate and viable, along with the limitations of implementation. Such distinctions may be made in terms of the content of training, types of work environments, and the levels of the jobs to which training is applied.

Types of content

Developing reference tasks, scaled mastery values, and repeated measurement is likely to be easier, and more generically applicable, for basic knowledge and skills. Bunderson, for example, notes the importance of using underlying cognitive constructs to validate representations of a knowledge or skill domain. Underlying scales of proficiency will involve dimensions of basic knowledge. Indeed, Bunderson's example of scaled mastery relies on a construct of basic adult literacy.

Thus, in the work setting, the content grist for the mastery measurement mill is likely to be categories of fundamental knowledge—for example, quantitative skills, verbal skills, technology literacy, problem solving, and similar domains. Emphasizing those fundamentals would be consistent with industry's growing recognition of the need for basic skills training and education to adapt to new technological environments (Jacobs, 1987). Mastery measurement would find a more comfortable and functional home with underlying domains like computer programming than with specific skills like being able to run a spreadsheet program. Specific skills, if they can help calibrate scales of mastery, can become reference tasks for the basic domains.

Types of work environments

A measurement science system can work effectively in a number of environments, but it may be used best where an integrated approach to job performance, learning, immediate feedback, and reinforcement are necessary. Highly automated process-oriented industries and industries employing computer-integrated manufacturing are likely sites. These environments allow the opportunity for creating a comprehensive and integrated performance system and also seem to be plagued with the constraints that make traditional classroom training difficult to administer.

Jobs that require a high level of knowledge or skill are another possibility. If worker turnover is high, or job rotation is required, a measurement science intervention may also be successful. Certainly, anywhere the consequence of poor performance is severe, a measurement science system should be considered. Such a system will likely work best in an environment where the content base is closely tied to a criterion of performance. Hence a training versus an education setting may be more appropriate.

To what environment is a measurement-science system least suited? If the content base is rapidly and justifiably changing, or experimental, a measurement science system may not be the best alternative. A measurement science system could be used in such settings, but it would will likely be costly. Of course, in a manufacturing facility with massive production investment or health and safety implications, the expense may be warranted.

Levels of jobs

Conceptually, a training measurement system should be applicable across hierarchical levels and functions in an organization. Practically, there may be a dilemma between where the approach is most needed and where it is easiest to develop. If the focus is on basic underlying mastery for decision-making, multiple-task application, or "cognitively complex' skill domains, high-level managerial jobs seem the best opportunity for application. Yet needs for technical training and measurable success are generally more keenly felt at the level of workstation operator or individual contributor. The design of a training and measurement system for the latter level should incorporate mastery requirements for a variety of job assignments and career mobility within the organization, rather than for particular jobs.

Monitoring the Measurement System

Once installed and adopted, a measurement system cannot be left on its own. Someone must monitor the system for accuracy, appropriateness, and overall problems, just as any other system is monitored.

Monitoring criteria

The monitoring and control procedures must be established when the system is designed. Several criteria should be considered in establishing these procedures, including the following:

- The system must be responsive to changes in the way particular tasks are completed. If new equipment, software, or procedures are installed, the skills, knowledge, and abilities required to complete tasks may change, and the mastery assessment system must change accordingly. The mastery map and reference tasks, training materials, feedback, and performance reporting all must be altered to reflect changes in the job.
- The system must be responsive to changes in job responsibilities. When changes create new tasks that are not included in the assessment system, management must first decide whether the new tasks should be included in the system and, if so, must arrange for new behavior analyses leading to changes in the system.
- The system must be responsive to changes in organizational structure or hierarchy. If new people will be receiving assessment reports, for example, management must be prepared to train them

in interpreting and using the reports. Newcomers to work groups must also be trained to understand and use the mastery assessment system.

- The system must be responsive to changes in performance standards. As Bunderson highlights, performance standards may increase because of improved training and measurement. They may also increase because of new equipment, incentive programs, or other environmental and motivational factors. Management should monitor these changes and, when the new performance levels are stable, alter the system to reflect those standards and to facilitate achieving them. If performance standards decrease, management should determine the cause of the change and take corrective steps. Possible causes of lower performance include problems with hiring practices or criteria, training, and appropriateness of the measurements. If the problem involves training or measurement, the assessment system must be reviewed.
- The system must be responsive to changes in the training environment. For example, if the mastery assessment system is computerized and new hardware is introduced, the assessment system may have to be revised to accommodate the change.

Monitoring techniques

It is not necessary to design elaborate techniques to monitor an assessment system. Many existing communication channels can be used to review the system and give indications of problem areas. A few channels are described here.

- Performance reports circulated to employees, supervisors, and managers may indicate significant changes in individual or group performance. Positive changes should be monitored and, if stable, included in the assessment system. Negative changes should be investigated and rectified.
- Trainees may be asked to evaluate the system during training and on the job. They can provide feedback about how helpful or threatening the system is and how they have been able to use the information provided.
- Supervisors and managers can provide feedback on how individual performers are using the system and how helpful it is in management functions. They can also report any discrepancies between formal reports and observations of workers.
- Management can review the system in light of organizational results. Performance indicators can be translated into monetary terms and compared with the rationale of the initial justification. Improved performance may not be attributable entirely to training and measurement systems, but management will often agree to

some percentage (say, 10 to 15 percent) that can be attributed directly to training. That percentage can be used to determine the amount of money saved or the return on investment.

Summary and Conclusion

Implementing measurement systems for training will be challenging for most organizations. However, when organizational change is planned and when barriers to implementation have been minimized, an integrated measurement system can enhance training and performance significantly. The key elements for an effective implementation include

- staged implementation, from awareness and front-end analysis, through adoption and evaluation;
- structured and systematic planning and justification of the system;
- calibration of tasks referenced to jobs, positions, or production systems;
- an organizational system for continuous measurement and feedback of progress toward mastery;
- formative evaluation of learners and the training process, coupled with summative evaluation of the impacts of training on the organization;
- a team approach to implementation;
- tangible incentives for participants;
- criteria and techniques for monitoring the measurement system.

Certain situations are likely to be more amenable to measurement systems than are others. A measurement system for mastery is likely to be more generically applicable to basic skills and knowledge than to specific work tasks. It may also be more appropriate in highly integrated manufacturing settings and in jobs requiring high skill levels than in situations where work content is rapidly changing. Finally, it may be most effective for managerial jobs or technical jobs in which mobility for the individual is expected. Taking advantage of those areas of opportunity can expedite the gradual absorption of measurement systems in industrial and business training. Even where implementation will be more difficult, it may still be appropriate.

References

Bowering, D.J., Fortune, J.C., & Wood, M.T. (1985). Validation as a formative process. In J.C. Fortune and Associates (Eds.), *Understanding testing in occupational licensing*. San Francisco: Jossey-Bass.

Dormant, D. (1986). The ABCDs of managing change. In *Introduction to performance technology.* Washington, DC: National Society for Performance and Instruction.

Georgenson, D.L. (1982). The problem of transfer calls for partnership. *Training & Development Journal, 36*(10), 75–78.

Gurnani, C. (Fall, 1984). Capital budgeting: Theory and practice. *Engineering Economist, 30*(1), 19–46.

Jacobs, J. (1987). *Training and public policy* (Technical Report). Ann Arbor, MI: Industrial Technology Institute.

Martelli, J.T. (1987). Management training: What's it worth? A performance-based evaluation. *Performance and Instruction, 26*(9)10, 32–36.

Mosier, N.R. (1986). *Financial analysis techniques: A review of the methods and their application to employee training* (Project No. 9). St. Paul, MN: University of Minnesota, Training and Development Research Center.

Shikiar, R., Wood, M.T., & Saari, L.M. (1985). *Training strategies for new technologies* (Technical Report to the Ford Motor Co.). Seattle, WA: Battelle Human Affairs Research Center.

Vasilash, G.S. (1982, October). Playing for high stakes in high technology. *Production.*

Wang Laboratories (1985). *Issues in information processing: Cost justification.* Lowell, MA: Wang Laboratories.

Wood, M.T. (1988). Training and education for advanced manufacturing technology. *Yankee Conveyor.*

Wood, M.T., Klimoski, R.J., & Del Gaizo, E. (1974). *Social incentives for technical training* (Technical Report to the U.S. Air Force Human Resources Laboratory, Report No. AFHRL74-10). Columbus, OH: Ohio State University.

Wood, M.T., & Stoll, H.W. (1988). *Needs for training in design for manufacturability* (Technical Report). Ann Arbor, MI: Industrial Technology Institute.

SECTION 4

Implementation of HRD Measurement

Alternative Interactive Instruction and Measurement

Scott A. Sayre
Director
Sayre Media Consultants

In many training environments, intelligent interactive instructional measurement systems are the best solution for encouraging and monitoring trainee progress. Bunderson's example in this monograph, of intelligent interactive video is an excellent description of what is now at the forefront of measurement science and instructional technology. Most trainers currently cannot afford to produce or adopt this type of system, however, because it requires large commitments of time and capital. Consequently, those trainers often continue to use the traditional lecture approach and some form of paper-and-pencil measurement. Nevertheless, there are many other opportunities for compromise between these optimal and traditional extremes. There are alternative types of instructional systems that offer trainers affordable and efficient means of interactive instruction and measurement. The focus of this chapter is on identifying elements from Bunderson's example that can be practically applied in both small and large training organizations.

The interactive video (IAV) instructional system in Bunderson's example has many advantages for both trainer and trainee. For the trainer, this system offers the ability to automatically cater instruction to individual needs while constantly measuring the learner's progress and mastery. This automated system also allows the trainer to integrate a variety of visual information and animation to simulate specialized learning environments. From the trainee's standpoint, this system offers a flexible learning environment in which he or she can interact fully with the training material, using a variety of input devices. In addition, because measurement is recorded formatively during each interaction, it is virtually non-intimidating and invisible. In Bunderson's example, a sophisticated, intelligent, interactive video system provides these capabilities. However other more accessible instructional systems offer many of these capabilities at only a fraction of the cost. These powerful systems are a new generation of what has been called computer-based instruction (CBI).

Early CBI systems had a reputation for being tedious, text-based systems, but they have come a long way in the last few years with the prolific influx of powerful, user-friendly microcomputer hardware and software. CBI systems now allow instructional designers with little or no formal knowledge of computer programming to produce highly interactive instruction and measurement very efficiently. Apple Computer's "Hypercard" (Atkinson, 1987) and Authorware's "Course of Action" (1987) are two excellent examples of new software packages that drastically simplify the steps involved in programming, giving even non-programmers the ability to produce the programs of which they previously had only dreamed (Goodman, 1987). Currently, both packages are available for Apple's Macintosh computer. Although programs designed with either authoring system may be used to control an external videodisc player (creating an interactive video system), this is far from necessary because the real advantage of these programs lies in the capabilities of the software and computer. These capabilities include various levels of interactivity, systems for embedding formative measurement, graphic-based simulations, and user-friendly operations.

Interactivity

CBI can be used in a wide variety of instructional modes, including tutorial, drill and practice, and simulation (Gagne, Wagner & Rojas, 1981). In all of these modes, interactivity is a key component in giving both the user and the instructional designer various elements of control. Interactivity allows the user to control progress through core instruction, to obtain supplementary instruction such as help or remedial information, or to review previously presented instruction (Schweir, 1987). The ways in which interactivity is used may differ greatly, depending on the objectives of the particular program. Interactive responses can range from keying a simple response to assembling and operating a computer-generated model. Instructors also benefit from user interactivity by having the computer automatically manage and evaluate student progress. Instructional designs may be adapted to the individual user's level and learning style (Tobias, 1987). This type of design interaction, along with content interaction such as problem-solving activities, can be recorded and analyzed by the program, using a variety of measurement techniques. The instructor then can use these records to assess an individual's progress and identify special needs.

Embedding Measurement

In most traditional forms of instructional measurement, the points of evaluation are very obvious and often stressful for the user. Also, formative evaluations usually are done in incremental chunks rather than continuously. Like IAV systems, however, CBI's interactivity allows the designer to monitor an individual's performance unobtrusively in a wide variety of ways. In addition to being able to support what Bunderson defines as first-generation computerized tests and second-generation computerized adaptive tests, CBI is also capable of third-generation continuous measurement. Each time the user interacts with the instruction, his or her input may be referenced on a specific criterion scale. In some cases, the content of a user's input may serve as the criterion, whereas in others, an indirect variable, such as time spent on a task or the number of correct responses, may be used. CBI's ability to conduct continuous performance measurement is not surprising because it is the microcomputer that performs the same functions in IAV.

Bunderson also defines a fourth generation of intelligent measurement systems capable of performing expert analyses of measurement data. Although the previously described CBI software is not capable of performing easily at this level, this advanced technology may soon be more accessible. Even IAV software is currently very limited in its ability to perform truly intelligent functions. As Schweir (1987) explained, "The research in this field is embryonic, yet it suggests new ways of thinking about interactive video production which will likely influence future generations of software" (p. 73). The same could be said for intelligent CBI.

Graphics and Simulation

Although current CBI systems are not capable of displaying video sequences without an external videodisc player, they can display and animate a wide variety of computer-generated graphic environments. Drawing and painting tools are available within these programs, and photographs and illustrations can be imported or digitized directly into the computer for use within an instructional program. Because of CBI's limited audio and its inability to reproduce video sequences, it typically is not considered an efficient medium for conducting simulations (Reeves, 1986). However, new CBI software allows the instructional designer to build sophisticated, user-manipulable animations. New audio-digitizing (recording) software and audible text-reading programs may also be incorporated within programs to expand their previously limited audio capabilities. Although these audio and graphic capabilities are still quite limited compared with Bunderson's videodisc example, CBI has grown up and beyond its purely text-based ancestry.

Trainee Friendliness

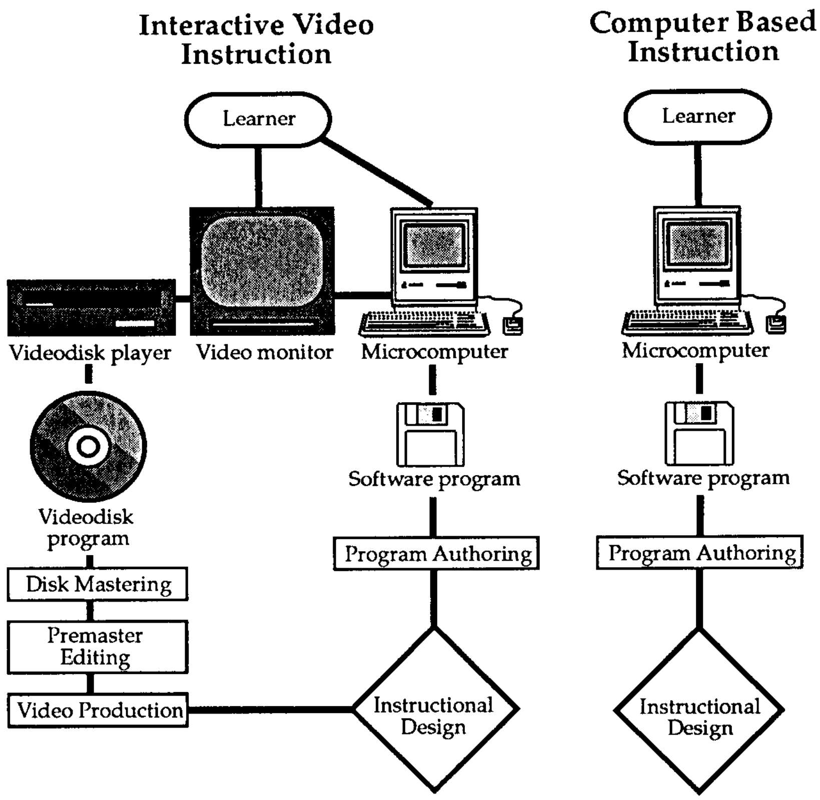
Like IAV, CBI can reduce computer intimidation of trainees, as well as increase interactive possibilities by offering alternatives to standard keyboard-based interaction. Depending on the instructional application, designers can choose the input device that best fits both the learner's characteristics and the instructional objectives. For situations in which the user has special needs and is incapable of using a keyboard, simplified input devices such as a mouse or a touch-sensitive screen may be used. More sophisticated training environments, such as engineering, may require the user to create or manipulate objects using an electronic pen. By integrating these input devices with unobstrusive measurement techniques in a graphic-based learning environment, trainers now can afford to apply instructional techniques that previously were accessible only with IAV.

Development Considerations

A CBI system has fewer components than an IAV system (see Figure 1). The comparably lower capital investment in both hardware and software can make CBI a viable solution in many instructional settings in which IAV would be too expensive.

Trainers must also consider the costs of program development in terms of time and personnel. The added components and the complexity of IAV increase the responsibility of the trainer. Although trainers may find the time to develop CBI programs themselves, it is often more economical for a trainer to play the role of content and measurement expert and have a staff member or freelance designer do the actual development. With CBI, this can be done easily by one individual rather than by the team required in the production of IAV programs (Kearsey & Frost, 1985). CBI avoids the media management problems of IAV and the complexities of standard programming languages, so trainers can concentrate on the design of instructional technology rather than on the tools to create it.

Figure 1—Basic system requirements



Conclusion

Instructional technology, like other high-tech fields, is rapidly advancing. With these advances come not only new products but also improvements in existing ones. To produce the most effective solutions to instructional problems, practitioners must continuously update their knowledge of products. Although many of the emerging technologies and techniques may seem intimidating to even the most competent practitioner, the only real threat is to those who ignore them. By continuously sampling and evaluating these ever-changing resources, practitioners optimize their current efforts and continue to grow with the changes. Those who rely on outdated information will have to absorb the costs.

References

Atkinson, W. (1987). *Hypercard* [Computer program]. Cupertino, CA: Apple Computer, Inc.

Course of Action [Computer program]. Minneapolis, MN: (1987). Authorware, Inc.

Gagne, R.M., Wagner, W., & Rojas, A. (1981). Planning and authoring computer-assisted lessons. *Educational Technology, 21*(9), 17–26.

Goodman, D. (1987). *The complete Hypercard handbook*. New York: Bantam Books.

Kearsey, G.P., & Frost, J. (1985). Design factors for successful video-disc-based instruction. *Educational Technology, 25*(3), 7–13.

Reeves, T.C. (1986). Research and evaluation models for the study of interactive video. *Journal of Computer-Based Instruction, 13*(4), 102–106.

Schweir, R. (1987). *Interactive video.* Englewood Cliffs, NJ: Educational Technology Publications.

Tobias, S. (1987). Learner characteristics. In R.M. Gagne (Ed.), *Instructional technology: Foundations*, (pp. 207–232.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Test Anxiety: Some Practical Advice

Gene W. Poor
Associate Professor
Bowling Green State University

Brian P. Murphy
President
The HRD Department, Inc.

The purpose of this chapter is to discuss a major problem with measurement, test anxiety, and to present some practical advice on how to solve the problem.

In one way or another, learners have always been evaluated to determine whether they have mastered the knowledge or skill required of them. Even members of primitive tribes were tested on heroism, skill in hunting, and knowledge of folklore as part of their initiation ceremonies (West et al., 1959). Today, many businesses and industries find they must use some form of evaluation to determine a learner's performance ability and growth.

But measurement doesn't have a good reputation among most learners, and anxiety about tests is a major problem for a large number of test takers. For example, a survey of 221 learners revealed that nearly 70 percent experienced some form of anxiety-related problems, both in the months before and during the examination (Davies, 1986). Symptoms included poor concentration, persistent worries, panic reactions, and minor health problems.

According to Salame (1984), most test-anxious learners have poor study habits and negative attitudes concerning their ability. Unfortunately for the test-anxious trainee, the current attitude of trainers is that examination stress is inevitable. Measurement is viewed as an integral part of learning in business and industry, and society expects test takers to cope with examination pressures.

Early research studies by Mandler and Sarason (1952) indicated that test anxiety invariably resulted in deterioration of performance on a test. In 84 more recent studies reviewed by Tyron (1980), the finding was confirmed. Tyron argued that highly anxious people worry extensively during examinations, thereby impairing their concentration and performance. Hembree (1988) also concluded that test anxiety causes poor performance and that anxiety relates inversely to self-esteem and directly to defensiveness and fears of negative evaluation.

The message of Bunderson's "Measurement Science and Training" is that the nation's work groups can attain high levels of expertise through the "application of deeper and more disciplined uses of science and technology and their handmaiden, measurement." If Bunderson's idea is true that the consequences of training are increasing, then HRD and measurement professionals need to directly address and resolve the problem of test anxiety.

Following are some practical suggestions for reducing the anxiety of both test takers and test makers.

Simplify the Terminology of Measurement

A cursory review of the measurement literature illustrates that the terminology is overwhelming, conflicting, and confusing. There is little wonder that HRD practitioners have test-maker anxiety. With all its acronyms, buzzwords, and statistical overtones, measurement appears unmanageable.

Illustration: An HRD manager continually talked about "cognitive test results" to company executives. When he used these words, the executives did not understand what he meant, and he did not clarify. Who was the loser? If the HRD manager had referred to "paper and pencil tests" or "knowledge tests," he would have kept the attention of his listeners. Measurement purists would argue that knowledge is the lowest level of cognition, but this level of fancy wordsmithing can only cause problems.

Illustration: A company that conducted middle-management training had an elaborate questionnaire to evaluate each session. Although the instrument appeared sophisticated, participants found it tedious and boring and the terminology confusing. When a brief response sheet that simply asked for the best and worst aspects of the program was adopted, participation dramatically increased. Managers gave valid and realistic reports of how the program benefited them and how it could be improved.

Further Simplify the Steps for Developing Measurement Instruments

In this monograph, Suzanne Saxe proposes the notion of "the one minute measurer" (in the spirit of the oversimplified *The One-Minute Manager*). Cheers to the spirit of the idea! In pursuing simplicity, HRD practitioners must first be able to develop measurement instruments that do what they are intended to do—and do it painlessly.

Illustration: A major corporation hired a professional consultant to assist with establishing test guidelines for evaluating management development programs. The director of the executive program described his exasperation with the consultant's report. The report, he said, was no more than "a review of the literature. I needed priorities and specific procedures to keep us from making big mistakes and to make improvements."

Illustration: A tool and die shop decertified itself from union status and became part of another personnel system within the company. Employees were then asked to define the structure, scope, and sequence of skills required for each occupational area. They determined how and when employee performance would be measured. A high level of acceptance resulted because the employees could control the transition into the new system. The simple and direct process proved successful enough to be extended to other parts of the company.

Stop Making Tests the Reason to Learn

All too often, tests are used as a negative reinforcement technique. The assumption is that learners won't learn unless given a reason, even if it is a bad one. In fact, however, adults in the workplace seldom learn simply to pass a test. This kind of negative reinforcement can only add fuel to the "test anxiety" fire.

Fortunately, most training programs are based on a rigorous analysis of expected work performance. Furthermore, most training programs expect participants to exhibit learning in a way that matches job performance. HRD specialists who take advantage of this relationship to job performance will not likely make tests that are "ends" unto themselves. The job will become the test, and, consequently, both the job and the test will become more credible to the participants.

Illustration: A four-year apprenticeship program included far more examinations than were required to determine apprentices' skills. The program was shortened, and participants were required to complete a concise series of final master projects. This change allowed participants to progress more rapidly and at their own pace. Trainees could devote attention to learning because they knew fully what skill they had to demonstrate at the end of each apprenticeship segment.

Put the "Fun Factor" into the Learning Process

Stamp out those HRD practitioners who don't design creative learning situations that are pleasant and enjoyable! Learning need not be painful just because the content is "meaty" and seems difficult to present.

By emphasizing formative tests and evaluations rather than traditional summative tests and evaluation, the opportunity for fun will increase. Furthermore, being compared with a standard is far less painful than being compared directly with other people. Given these two conditions, evaluation can celebrate accomplishments rather than shortfalls. Toss a few humorous questions or humorous tasks (ones that won't count against anyone) into the measurement device, and, who knows, there may even be a few belly laughs from some of the participants.

Illustration: Instructors at an electrical utility company needed to meet certification requirements to satisfy government regulations. Third-party evaluators contracted to evaluate instructors' performance in a variety of training situations. The evaluators increased the "fun factor" by visiting with the instructors prior to the evaluation. Then, during the evaluation, the evaluators actually participated in the training. Attempting tasks such as welding and fire fighting made the evaluators' jobs more enjoyable. Also, because the evaluators fit into the classes, the instructors were noticeably more comfortable.

Search for Unusual Ways to Measure Learning

Standard testing formats can present cues that raise test anxiety. Therefore, HRD professionals should use measurement methods that do not present these cues. Activities can be created that will evaluate a learner's knowledge and performance, yet hide the evaluation itself. Simulation, role playing, and games are some examples of activities that can be designed to train people, to have them practice skills, and to measure their abilities simultaneously.

Illustration: Supervisors in a high-tech factory were asked to develop expanded supervisory skills. All the supervisors knew that the company's new emphasis on employee participation meant they needed to demonstrate new behaviors. Measurement occurred during supervisory development sessions that required specific actions on their part to demonstrate their understanding of the new concepts. To the surprise of the supervisors, they were encouraged to create their own situations to practice. The increased reality of the practice sessions helped them focus on accomplishing a particular goal using the new skills. After the sessions, the supervisors remarked that they didn't notice the evaluation built into the learning process.

Build Continuous In-Process Measures into Training

One of the things that makes measurement so anxiety raising is that the test becomes a focal point for determining personal experience or knowledge. If measurement is a continuous part of the learning, it can disappear into the learning method.

Bunderson's proposal allows measurement to be embedded into the instructional design and instructional technology. By not "stopping to take a test" and by allowing the natural flow of learning to occur, the forms of measurement are potentially as varied as the content being taught. Furthermore, graphics and gaming possibilities of the interactive technologies offer new options for measurement.

Illustration: A manufacturer put measurement of learning in the hands of the trainees. They chose on-the-job quality and quantity product criteria and the continuous measurement of their own product quality and quantity as the focal points of their formative learning discussions with their trainers.

Conclusion

Taking the anxiety out of measurement will be a slow process, at best. It has taken a long time to move from primitive tests of bravery to the present sophisticated, standardized tests. The reality is that many people think measurement is a four-letter word, spelled T–E–S–T! As one disenchanted learner said, "Every time I know the answer, you change the question!" This statement is both humorous and sad. Unfortunately, it sums up the popular perception that measurement is unfair.

The question remains, "What can you do about test-taker and test-maker anxiety?" We have offered some practical and specific advice. The big issues—(1) ensuring test validity, (2) having the value gained from the measurement exceed the cost of measuring, and (3) ensuring ease of implementation—can never be ignored.

It is important to recognize that most of the advice in this chapter is embraced in Bunderson's proposals:

- to move from an emphasis on summative evaluation to continuous formative evaluation;
- to embed measurement in training so that it never looks like testing; and
- to make the fidelity of the test high so that "being able to do the work" is the real test.

Bunderson does not propose making measurement fun, although he certainly doesn't exclude the possibility. We think a little fun in measurement is still possible. Hey, did you hear the one about. . . ?

References

Davis, D. (1986). *Maximizing examination performance.* New York: Kogan Press.

Hembree, R. (1988). Correlates, causes, effects, and treatments of test anxiety. *Review of Educational Research, 58*(1), 47–77.

Mandler, G., & Sarason, S. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology, 47,* 11–173.
Salame, R. (1984). Test anxiety: Its determinants, manifestations, and consequences. In H. VanderPloeg & C. Spielberger (Eds.), *Test Anxiety Research* (pp. 83–121). Hillsdale, NJ: Erlbaum.
Tyron, G. (1980). The measurement and treatment of test anxiety. *Review of Educational Research, 50*(2), 343–372.
West, P., Pinter, R., Ryan, J., Aleck, A., Crow, L., & Smith, S. (1959). *Educational Psychology.* New York: Barnes and Noble.

One-Minute Measurement

Susanne Saxe
Manager, Management and Employee Development
Central Training Group
Wells Fargo Bank & Co.

HRD professionals are well aware that measurement science is a piece of the much larger picture of improving human performance. As Bunderson points out in the chapter "Measurement Science and Training," however, the application of measurement science to the development of human competence over time has been minimal at best in HRD and has been focused primarily in the diagnostic or placement area. The possibilities for improving this situation are astonishing and exciting; new mastery systems for developing human competence will be available by the year 2000. Early prototypes are already emerging from the fields of performance engineering, artificial intelligence, and computer science. The purpose of this chapter is to suggest a first step that will increase managers' effectiveness in using measurement science and that also will be aligned with today's fast-paced, bottom-line-oriented environment and the goal of moving employees toward increased competence.

The Need for Change

Traditionally, HRD departments assist employees in improving their performance by offering them standardized basic training on an established schedule. Employees are then let loose, with the hope that the training will improve their performance and that supervisors will reinforce training back on the job. To ensure improved performance, Bunderson proposes the use of mastery maps with domains defined in terms of knowledge elements, job elements, and types of tasks performed by novices, average employees, and masters. Mastery maps are flexible measurement tools that aid learning and instruction. Employees may follow various paths, skipping certain domains and focusing more heavily on others, as they move toward the goal of attaining exceptional levels of competence.

Bunderson proposes that the implementation of training, especially measurement systems, is critical to the success of developing human competence. If the primary goal of the HRD community is to improve human performance and to use new technologies and refine current methods, then HRD professionals can no longer afford not to measure levels of competence.

The Role of HRD

Most HRD professionals evaluate training primarily for the purpose of determining if workshops, self-studies, and seminars have met the stated course objectives. Not many practitioners, however, have systems in place for measuring steps to mastery. HRD professionals in business and industrial environments often find that measuring the steps to mastery is neglected for the practical reason that the reward is not usually worth the effort. As Michael LeBoeuf (1985) says, "We get the results that we reward." It is the role of HRD professionals to educate managers about the value of measuring increased competence and to show that measuring the increasing competence of staff is a way to achieve business results. It is also the responsibility of HRD professionals to provide workable systems and methods that can be used easily to collect data, develop employees, and measure results—all for the purpose of improving performance and thereby contributing to the company's success.

This chapter presents a model for "One-Minute Measurement." The process is guided by the climate of today's organizations, which are market driven and emphasize immediate results, quality and service, simplicity, flexibility, and especially the importance of the bottom line. One-minute measurement allows managers to focus on a few key questions and use data to improve performance and increase employees' contribution to the company's profitability. One-minute measurement can serve as assurance for the HRD department in meeting the bottom-line requirements of the organization while pursuing best practices and innovations as prescribed by Bunderson.

The One-Minute Lifestyle

The One-Minute Manager by Blanchard and Johnson (1982), made the *New York Times* best-seller list for an entire year for a good reason: it addressed readers' apparent needs for quick and effective management techniques. It was simple and easy to read, and, according to the authors, it worked because it was based on sound behavioral and human relations theories.

If quick is a criterion in today's marketplace, instead of beating their heads against the wall trying to implement thorough evaluation/measurement systems, HRD professionals can borrow from *The One-Minute Manager* and persuade managers to use "one-minute measurement" techniques.

Getting managers to buy into the idea of using data for developing human performance is the first step. Most managers are adept at using financial, sales, and production data to analyze productivity and profitability. So, why don't they collect data to use in developing human performance as well? Managers who do collect data on employees usually do it for performance reviews, not for developing competence. Managers' perceptions of the value of measuring human competence is neutral at best and often negative. To counteract this perception, HRD professionals can coach managers on one-minute measurement and its value to individuals, the group, and the company's bottom line.

The One-Minute Measurement System

The One-Minute Measurement system comprises three phases, each of which can be defined by a question:

- Are you collecting data?
- What do the data have to do with the employees' and department's performance goals?
- How does current productivity compare with productivity last week, month, quarter, or year?

Figure 1 shows examples of how to implement each phase. The goal of the system is developing human performance to meet businesses' and the world's ever-changing needs. It does not deal with learning options that should be used, but rather with the measurement of competence.

Figure 1—The one-minute measurement implementation phases and methods

High Perceived Value of Measuring = What is Measured

PHASES	METHODS
I. COLLECTING DATA Are you collecting data?	• Use data currently available • Capture information in casual conversation • Examine feedback • Use simple measurement tools • Test one thing at a time
II. DETERMINING THE DATA'S USEFULNESS What does the data have to do with employee and department's goals?	• Ask: what is relevant? • Focus on the few critical issues
III. COMPARING THE COLLECTED DATA OVER TIME How does current productivity compare with productivity last week, month, quarter, or year?	• Examine data over time

Phase I: Collecting data

It takes less than one minute to ask, "Are you collecting data?" In many cases, data already being collected are not being used for the purpose of improving human performance or are being passed on verbally but not being captured for later use. For example, how many managers pass along positive feedback to an employee or supervisor and yet never write it down for tracking purposes? If, for instance, a manager finds that the time she spends rewriting an employee's report decreases after training from one hour to 10 minutes on average, she should write this information down for later use. It takes very little time to compile such data, which can be very useful in tracking performance.

The following are a variety of one-minute-measurement techniques for collecting data.

Use data currently available. This is an easy way to begin. See what data exist either on paper, on computer, or in the memories of employees. Determine if this information is useful and relates to the goals of the department.

Capture information in conversations. Casual conversations with peers, managers, subordinates, and customers can be a valuable source. Much information is exchanged on elevators, in hallways, on the phone, and in meetings that is not captured. Collecting information from such conversations can be simple. Managers can mark it in their calendars, have a sheet handy by their desks with employees' names on it, have a flipchart in their office to write on, or use a dictaphone or tape machine.

Examine feedback. Managers are aware of the strength of positive and constructive feedback for improving performance. Many use it as part of their management style. Feedback is another opportunity for collecting data to develop human performance. Managers can ask themselves such questions as "How many times have I positively reinforced this employee for making a difference this month?" and "How often do I wonder if this person is on track?" Measuring the frequency of positive, constructive feedback may help managers determine how well the employee is performing.

Use simple measurement tools. There are a wide range of tools available, ranging from simple to complex. Swanson and Sleezer (1987, p. 10) described a two-question instrument for evaluating training that would be useful for one-minute measurement (see Figure 2).

Figure 2—Sample questions for determining satisfaction with training and development activities

	Strongly Agree	Agree	Disagree	Strongly Disagree
1. Employees have performed better at their old job or have been able to perform a new job following training.	______	______	______	______
2. Attending the training was a good use of the employee's time.	______	______	______	______
Additional comments:				

Data on performance collected during training sessions can be passed on to managers to emphasize reinforcement of skills learned back on the job. An example of this would be collecting data on the final skills practice feedback form in a skills training workshop and providing it to managers for performance comparisons before developmental activities (training, job aids, OJT, etc.) and after. This is especially easy to do if the group or organization regularly collects the data. For example, data on sales volume before and after a sales training program can be compared. Swanson and Sleezer (1987) also discuss the use of cost-benefit analysis to determine the economic value of training or the effect of developing human performance on the bottom line.

Other measurement tools can include customer satisfaction ratings for training received, as well as assessments of levels of increasing competence. Employees can ask themselves the following:

- In what areas of knowledge and skill do I need help?
- How will I obtain the knowledge and skills I need?
- Now that I have the knowledge and skills, what am I doing differently that demonstrates increased competence?

Self-assessment tools are valuable because they base the assessment on the widest possible range of observations. In addition, they can be used to determine how an employee perceives competence or improved performance on a specific skill or project. As Brown (1983) stated, "The basic assumption behind this approach is quite simple: The individual is in the best position to observe, describe, and report upon his or her own behavior. After all, you are the only person that is within your skin 24 hours a day, 365 days of the year; you are always present to observe your own behavior" (p. 360).

Test one thing at a time. When asked the question, "Are you collecting data?" managers may give a variety of responses. The One-Minute Measurement technique helps managers focus by suggesting they test one performance dimension at a time, collect data on one performance issue, and focus on improving that specific dimension. This simplifies the activity and helps managers to focus on critical performance areas.

Phase II: Determining the data's usefulness:

So the managers have collected the data—now what? The second question addresses its usefulness. What do the data have to do with the employer's, department's, or employees' performance goals? The One-Minute Measurement technique asks, "What data are relevant to improving human performance for the purpose of meeting or exceeding our department's goals?" It is helpful to focus on the one or two key factors that currently contribute to or interfere with meeting the goals. In a collective sense, the ability to focus on improving the critical skills needed in a business can make or break a company over the long haul.

Phase III: Comparing data collected over time

Data collected over time will provide a wealth of information on how human competence increases or doesn't increase over time. How does current productivity compare with productivity last week, month, quarter, or year? Managers and HRD professionals can compare data on the performance of individuals and the group to determine which alternative learning activities are needed to develop mastery and which activities have resulted in improved performance. Data expressed in charts, numbers, or graphs can help managers quantify the perceived value of developing employees.

Conclusion

Measurement science is an important field that needs to be better understood and implemented than it has been in the past. The first step is achieving management buy-in of its value, cost, ease, and fit with the organization's fast-paced business and educational climate. The One-Minute Measurement system is a practical method for implementing measurement science in the workplace. It is a beginning, one minute at a time, aimed at measuring one variable and tied to the bottom-line impact. It is also a step toward instituting more widespread and effective use of measurement science principles in today's organizations to prepare American business and industry for mastery systems by the year 2000.

References

Blanchard, K., & Johnson, S. (1982). *The one-minute manager.* New York: Berkeley Books.

Brown, F. (1983). *Principles of educational and psychological testing.* New York: Holt, Rinehart, and Winston.

LeBoeuf, M. (1985). *The greatest management principle in the world.* New York: Berkeley Books.

Swanson, R., & Sleezer, C. (1987). Training effectiveness evaluation. *Journal of European Industrial Training, 11*(4), 7–16.

Issues for Professional Practice

Dale C. Brandenburg
Associate Head, Division of Measurement and Evaluation
Office of Instructional and Management Services
University of Illinois

Bunderson's thought-provoking chapter, "Measurement Science and Training," raises a number of issues concerning the application of measurement to HRD practices. To examine the general concern about who is likely to carry his ideas to fruition, three questions will guide the discussion:

- What is the current level of measurement expertise in the HRD profession?
- How can current measurement practices be upgraded to implement Bunderson's ideas?
- What is a feasible way to start the process of upgrading measurement practices?

The focus here is upon those professionals within the HRD community who will conscientiously consider applying Bunderson's concepts to the HRD process.

Measurement Expertise and HRD

Current status of HRD professional staff

Typically, two groups of HRD professionals are in position to apply measurement science. One group is the instructional design/development staff and the other is HRD managers. This section address the first group—the staff—and is directed toward their use of measurement in the typical classroom training situation where measurement of individual learning is emphasized.

A competent instructional designer can analyze content, specify a content domain, break that domain into terminal objectives related to the business problem at hand, articulate lower-level objectives, and sequence the content for learning. The result is a plan or instructional sequence for application to the business problem.

As a result of their efforts, competent designers/developers can produce effective courses and programs. If these developers are also skilled in criterion-referenced instruction or competency-based instruction, it is likely that their courses will contain questions and exercises designed to check on skill development. In general, applied measurement will be consistent with the type of skill taught. For example, process-oriented learning objectives (working as a team, techniques of conflict management, consensus decision-making) call for involvement methods, such as case studies or role plays, in which checklists are likely to be used as evaluation tools. Knowledge acquisition usually calls for lecture or self-paced material and objective tests, such as multiple choice, are more likely to be used to evaluate learning.

In many HRD staff positions, however, no distinct differentiation is made among needs analyst, designer, developer, instructor, and evaluator. One person may fill all these roles. HRD staff members with such broad responsibilities are not likely to have had a thorough grounding in classical test construction techniques or in test theory. Many do not know how to apply basic concepts to the construction of test items. Because of their lack of expertise, such individuals produce test items that look like "add-ons" or afterthoughts. Their test items often measure simple recall of knowledge. For measurement of process-related skills, they give little thought to quantifying measures of embedded activities (case studies, role plays), which could yield important information about individual learning.

To illustrate the ineffective testing of a knowledge-acquisition objective, let us examine a test item written by an inexpert instructional designer. The following item was written for a test in a training course for electronic technicians:

"Motorboating" is often due to ____________________.

a. reduced filter capacitance.
b. open rectifier diode
c. bad tube
d. volume too loud
e. none of these

Note that there are two ways to obtain the correct answer to this item without knowing the content. The first is to see that the correct answer is the only response with a period after it. The second is to realize that the correct response is the only one that makes grammatical sense. If this item had been written for other than a single classroom learning situation, its obvious problems would have been corrected.

Another example shows how a poorly constructed test item from a training program on statistical quality control could be revised by following some reasonable rules for test construction.

Here is the original item:

A process design team has diligently used a common part in order to reduce inventory. This part is used in 10 different products. The materials people have found it difficult to track which product is using what quantity of parts. To control this, they have assigned separate part numbers to the common part based on the product in which the common part is used.
This means:

a. There is no change in inventory because of this policy. A common part is a common part, no matter what it is called.
b. The common part is now functionally 10 parts because with the separate part numbers each must be handled independently and now has about 10 times the overhead compared to when it was a single part number.
c. Inventory goes down because of the greater tracking of the parts.

Here is the revised item:

A part is used in 10 different products. The materials people have assigned separate part numbers to the common part based on which product makes use of the common part. What effect does this assignment have on inventory overhead?

a. increases
b. decreases
c. remains the same

This item was revised by eliminating irrelevant and possibly misleading information from the stem, or main part, of the item. In general, test item stems should provide only the information needed unless the additional objective of filtering out unnecessary information is also being tested. In addition, the items should not teach as well as test. The revised item above gets directly to the point.

In summary, the lack of background in classical test construction techniques and test theory on the part of many designers/developers can result in poor measurement of individual learning in HRD training programs.

Current status of HRD managers

Current studies (Brandenburg & Schulz, 1988) show that in recent years there has been (a) a general increase in the level of sophistication in training evaluation practice, (b) a greater demand for high quality internal training programs, and (c) an increase in the types and comprehensiveness of training evaluation methodology. The use of standard (classical) testing procedures has gained more attention as managers have attempted to improve all phases of the training process. (See Shrock & Coscarrelli, 1986, for a discussion of competent current practice.) Even with these gains, however, there appear to be great deficiencies in reported practice (Parker, 1985).

Although most training managers may not be expected to know measurement theory in depth, they should recognize good practice. However, misconceptions among training managers regarding measurement practice commonly exist. Two major reasons for this situation are that

- the concepts of measurement science are often heavily laden with sophisticated mathematical jargon, and
- learning how to apply current measurement technology requires considerable practice and experience.

To the purist, eliminating the mathematics dilutes the value of the theory. To the practitioner, however, using the sophisticated jargon and technology diminishes the utility of application. Thus, the image of testing is *not* enhanced through the use of complex measurement theories and formulas. For example, the assumptions required for using latent trait analysis may be rejected by HRD staff members. Bunderson's paper does not address these very real barriers to implementation.

Doing a good job of measurement requires an intimate knowledge of the content, the patience to write a test item, and piloting the first drafts.

Also, the requirement for knowledge of the content to be tested means that either the trainers must gain the appropriate knowledge or they must work with a subject matter expert who has the knowledge. Either way is time consuming.

The amount of time required to do the measurement job right deserves specific comment. An experienced test writer composing a typical objective, multiple-choice item and response for a nationally standardized test might spend 30 to 45 minutes on a comprehension-level item and 60 to 90 minutes on an application-level item. In a typical training situation, the time needed to write such items would be approximately one-third of that above, if the writer were experienced at the craft. Given these time requirements, it is no wonder that high-quality, well-received tests are not common in training environments.

In summary, barriers to implementing the competent practice of measurement science include a variety of constraints that need to be addressed.

New Emphases in Professional Strategies

HRD practitioners in the future will need knowledge and understanding of new measurement techniques and technologies. This calls for a new paradigm of measurement that incorporates assessment at the various organizational levels—the individual learner, the job, the production unit, and the organization. In short, this emphasizes a comprehensive, problem-oriented view for measurement whereby it can impact human performance problems and employee development in a variety of ways.

Doing better what we already know how to do

To fully develop and implement this new paradigm, prior to actually adopting Bunderson's ideas, HRD practitioners should seek to improve what they already know how to do in the area of measurement. In tracking either individual or group performance, they should progress professionally on a data-based path from novice to expert and use data-collection procedures that are business related.

Today's claim that measurement must be business related is parallel to the educationally related arguments advanced almost 40 years ago by one of the pioneers of objective achievement testing, E.F. Lindquist. Lindquist (1951) used the labels "immediate objectives" and "ultimate objectives" to differentiate aspects of instruction that are realized in the classroom from those not likely to be fully realized until long after instruction has concluded. This dichotomy is consistent with the HRD-related distinction between "knowing it" and "doing it." The measurement of immediate objectives in the classroom, according to Lindquist, is generally accomplished through "indirect" instrumentation. Lindquist cautioned educational practitioners against using indirect measures, such as multiple-choice tests, as surrogates for measures of real-life applications. HRD professionals often forget these roots. However, Bunderson's ideas provide a potential solution to such problems and thus should be recognized as important steps toward "direct" measurement of immediate objectives. (Lindquist would probably agree that we should have been at this point years ago.)

Concepts for professional advancement

Three concepts proposed by Bunderson gained considerable support among the participants at the symposium: continuous measurement, response tasks, and the integration of measurement science with instructional science. The reader will note that several chapters in this monograph refer to these topics or their derivatives. Consideration of the three concepts simultaneously results in "embedded" measurement, that is, measurement that is integrated with instruction and employee development and does not look like an afterthought.

Moving from current measurement practices that concentrate on "knowing it" to "doing it" seems a particularly powerful concept. Items currently used in testing have been criticized as offering too narrow a breakdown of learning. Disenchantment with present objective testing technology has drawn criticism, even from within the profession. Wilbert McKeachie (former president of the American Psychological Association) remarked at a recent lecture (1988), "Objective tests, mainly multiple-choice and true-false, are one of the greatest disservices psychology ever did for education."

When "doing it" response tasks are used as a testing tool, measurement looks more like the training it is intended to track. The content validity of the measurement can be upgraded in the same way that typical multiple-choice test items are upgraded. Using response tasks will have an important impact on the direct measurement of immediate objectives. Some use of response tasks is evident in current training practices where the goal is teaching/learning process skills. The integration of measurement with instruction has occurred in programs that use case studies, role plays, or simulations not only to teach and practice target content but also to measure results of the training. Bunderson proposes the calibration of such measurements to quantify, as well as describe, learning results. Although the validity of such calibration must be empirically justified, the concept shows promise for significant impact on future training design.

Measurement and its role in HRD

If we assume that greater measurement expertise on the part of HRD professionals is needed, we should also recognize that the role of such expertise is limited. Measurement as it is most commonly applied to evaluate training is tied to line and senior management's perception of the merit and worth of HRD interventions. As Bill Wiggenhorn (1984), director of Motorola's Training and Education Center, stated:

> In determining what works, perceived return is often more important than documented return. Perceived returns pertain to whether managers believe they got something of value or not. No matter what the numbers say, if they don't think they got value, they didn't get it (p. 13).

Wiggenhorn further stated that if HRD professionals do not understand what it is managers must produce, measurement technology will languish unused. Advanced measurement concepts can help HRD professionals demonstrate to managers how HRD helps solve their problems. Thus, the introduction of new tools for the measurement process must be viewed in terms of assisting in the attainment of business goals.

Expanding Professional Measurement Capabilities

In the near future, accurate measurement applications will become more important in the HRD process than they are today. The way to begin putting more measurement into HRD is to start with those processes under our management, such as applying quality control when pilot testing courses. HRD professionals should provide measurement results that reflect the quality of program design and development—not the ability of the learner.

Consistent with this description of the role of measurement is the fact that most organizations do not need to hire a full-time person with substantial measurement expertise. Examining the potential contribution of measurement can best be handled through the use of external consultants. Experienced consultants can provide ideas for incorporating measurement where needed, or they can assist in identifying those opportunities for measurement application that will reduce business risk in HRD operations. In addition, the use of an internal audit to assess measurement practices would be worthwhile. All levels within the HRD process could be examined—organization, business units, subsystems, jobs, and individuals—everywhere that the application of measurement could yield direct benefit. Such an audit should also attempt to identify existing expertise within the staff, which will provide linkage points for external consultants.

Conclusion

In response to Bunderson's chapter, I chose to focus on a number of issues related to the current practice of measurement science by HRD professionals. First, I pointed out that general evaluation methodology incorporates the application of sound measurement principles and ideas. Second, I described the current status of skills and knowledge among HRD staff and management and the needs for improvement in order to upgrade the practice of measurement. Third, I proposed that accurate measurement application to business problems will become more important to the HRD process. Finally, I suggested that more measurement be added to our work, starting with the processes most completely under our control—our own training programs. In summary, I believe we should move measurement practice to a higher level of quality by challenging or initiating the development of HRD staff.

References

Brandenburg, D.C., & Schulz, E. (1988, April). *The status of training evaluation: An update.* Paper presented at the National Society for Performance and Instruction Conference, Washington, DC.

Lindquist, E.F. (1951). Preliminary considerations in objective test construction. In E.F. Lindquist (Ed.), *Educational measurement* (pp. 119–158). Washington, DC: American Council on Education.

McKeachie, W. (1988, March). *Developing critical thinking in college instruction.* First Lilly Fellows Lecture, University of Illinois, Urbana, IL.

Parker, B.L. (1986). Summative evaluation in training and development. *Journal of Industrial Teacher Education, 23*(2), 29–55.

Shrock, S., & Coscarelli, B. (1986). Ringing out the bell-shaped curve: Preparing criterion-referenced tests, a four-part series. *Data Training, 5*(10), 48–51.

Wiggenhorn, W. (1984). Evaluation challenges in human resource development. In *Joint ventures between business and higher education: Human resource evaluation.* Evanston, IL: Arthur Anderson & Co. and Northwestern University.

Evaluation in Perspective

Sandra J. Merwin
President,
Sandra Merwin Enterprises

As the utilization of measurement is increased in organizations, it is important to focus not only on implementation but also on ethics. A code of ethics is simply a set of values-driven guidelines which govern the conduct concerning the data-gathering process and the way in which data will be used. A code of ethics for measurement includes a set of guidelines about such issues as confidentiality, what data will be gathered, the role of HRD/training, the power of inference, and how the data will be used.

Most of these issues can be addressed by focusing on the power inherent in measurement. Almost everyone is familiar with the axiom that knowledge is power. Today, if you have numbers to back you up, you have power. Those who do not have the ability to interpret the numbers may rightly fear how measurement data will be used. These fears can be categorized by three questions.

- How do we know we're measuring the right stuff?
- What do these data determine?
- Who determines what the data mean?

Measuring the Right Stuff

How do we know we are measuring the appropriate skill or competency? In his chapter, Bunderson paints a scenario of an enlightened management group. The first step these enlightened managers took was that they "clearly identified high levels of human expertise, with real people." Although we may have the capacity to identify the individuals who are experts, we must also use developed methods of breaking down their unique qualities and competencies into valid content for training and measurement (Swanson & Gradous, 1986).

In an actual situation with telephone operators, it was discovered that the competency items that were determined to qualify an operator as an expert were not accurate. Thus, the telephone training consisted of teaching the new operators inappropriate methods of dealing with problems.

The first question, then, remains crucial—"How do we know we are measuring the right stuff?" The second question—"What do the data determine?"—focuses on personal fears with job loss and failure.

Data: Power or Resource?

Bunderson defines summative and formative evaluation as follows: Summative evaluation is sought by policy makers "either as a way to prove that programs are achieving good results, or as a way to attack current policies." Formative evaluation means to use data to improve a system as it is being developed or used.

Many corporate employees who have been indoctrinated into the "knowledge is power" school of management do not understand the difference between summative and formative evaluation. All data are considered resources. These data resources are used to make budget decisions, hiring and firing decisions, and decisions about the competence of a department. Although the information may be gathered as "formative data," it is possible that it will be used to make key decisions. For example, consider an organization in which six service centers are trained in a new technology. Built into the training system is formative evaluation. At the end of the course, it is revealed indirectly that one center performs at a lower learning level than the other five. When budget resources necessitate a cutback of service employees, will this data be used to determine which department to cut? Employees may have good reasons to be suspicious of data which can be used as a basis to fire, demote, or punish them.

In addition to training data being used to make organizational decisions, data can be used inappropriately in other ways, as in the case of nonconfidentiality.

The Buckley Amendment and the courts have outlined the rights of students insofar as that information is concerned. It is imperative that instructors realize that the Buckley Amendment provides access to educational records for students at all levels (McAfee & Cote, 1985).

Consider the confidentiality issue in the context of the role of a training department. With measurement, training now can determine who is learning faster, who performs better in simulations, who is not performing according to company standards, and even who needs to be replaced.

What do you do when you discover that one supervisor has a much slower learning level? Do you put this data in his file? Will this data affect his next promotion? What if this supervisor is ranked by his employees as a top-performing supervisor? If you, as a trainer, discuss this information with a peer or the supervisor and casually remark that the supervisor is "a slow learner," you could end up in court for breach of confidentiality.

Who Determines What the Data Mean?

Given a set of data, different people may draw somewhat different conclusions. Yet the language usually used in reporting investigations makes the process of inference sound objective: "Statistics show that. . . ."

In the past, training departments have complained of their lack of power in their organizations. With measurement, training departments can gain power. Training departments also can use data to infer that learning programs should be continued or stopped. For the most part, organizations have used data as summative.

In order to move from a summative to a formative data position, new beliefs and values about measurement need to be fostered in organizations. An operating philosophy and a clear-cut set of values-driven guidelines will be helpful in countering the natural resistance to measurement and testing.

Although the testing and measurement may be unobtrusive, the fact the testing and measurement are taking place must be stated along with information about:

- how the data will be used;
- who has access to the data;
- the purpose for gathering the data;
- how employees are protected so the data will not be used to punish or disgrace them;
- how the data will be communicated to the employee, to the organization, and to the employee's supervisor.

Bunderson states that "a revitalized measurement science can be of significant help in achieving new levels of effectiveness." This is true, but only if the individuals in organizations develop a new understanding of measurement. When an organization chooses measurement as a method of achieving new levels of effectiveness, it must also look at the values, beliefs, and ethics of this change. The principles and guidelines in the following table are useful for using measurement in training.

The challenge is to handle the rewards and risks of measurement, while leading the organization and its leaders beyond summative data. It is not enough to trust an enlightened management group or to trust that everyone understands the difference between a summative and formative approach. Up-front agreement on measurement guidelines will keep the organization out of ethical trouble. Most important, each organization will need to determine how the formative data can empower the individual. As evaluation increases its role in HRD/training, practitioners must continually struggle to maintain a perspective of evaluation in the context of the organization's mission and the ethics of our society. The key opportunity for a revitalized measurement science is to use formative data to empower employees to improve their levels of performance and competence.

Table 1—Measurement principles and guidelines

	Accountability
Principle	• Acknowlege responsibility for using the highest measurement standards possible.
Guidelines	• Identify who is responsible for measurement. • Clarify the measurement. • Reveal the intended use of the measurement data. • Acknowledge the rights of participants, instructors, and others to question procedures or outcomes. • Make measurement data available to participants, their managers, and organizations.
	Confidentiality
Principle	• Acknowledge the individual's and organization's right to privacy.
Guidelines	• Inform individuals being measured of the purpose for which the data or information is used. This includes both processed information, such as scores based on reponses, and the raw data upon which the processed information is based. • Do not collect any critical information that cannot be protected adequately from improper disclosure.
	Accuracy
Principle	• Assure that measurement procedures are reported and results are accurate.
Guidelines	• Verify that measurement procedures conform to accepted standards. • Have individuals competent in measurement procedures review and assess the results. • Construct tests that are reliable for the intended use. • Adhere to professional measurement standards.
	Fairness
Principle	• Establish fair and proper procedures for using measurement information.
Guidelines	• Provide clear outcomes for using the evaluation or measurement procedure. • Caution individuals against drawing conclusions not supported by validity evidence. • Warn individuals about anticipated misuses of measurement data. • Encourage individuals to put measurement data in the appropriate perspective (i.e., post-reaction surveys reveal the satisfaction of workshop participants, but they do not reveal the amount of learning that took place). • Investigate all complaints about improper use of measurement data.

References

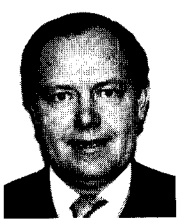
McAfee, J.K., & Cote, L.S. (1985). Avoid having your day in court. *Training & Development Journal, 39*(4), 56–60.
Swanson, R.A., & Gradous, D. (1986). *Performance at work.* New York: Wiley.

About the Authors

J. Diane Beezer-McClure
Branch Manager
Honeywell Protection Services Division
180 Clinton Avenue North
Rochester, NY 14604
716/546-7053

Diane Beezer-McClure is the Rochester branch manager for Honeywell Protection Services Division, serving upstate New York's business and home fire, burglar, and access control protection needs. During her 4½-year tenure with Honeywell, she has also been corporate manager of assessment and a defense systems division staffing and EEO manager.

Prior to joining Honeywell, Beezer-McClure studied counseling and education at The Pennsylvania State University, receiving her B.S. in 1978 and her master's degree in 1980. She then worked at the university as an admissions counselor.



Dale C. Brandenburg
Associate Head, Division of Measurement and Evaluation
Office of Instructional and Management Services
University of Illinois at Urbana-Champaign
307 Engineering Hall, 1308 West Green Street
Urbana, IL 61801
217/333-3490

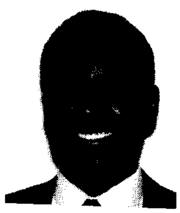
Dale Brandenburg is associate head of the Division of Measurement and Evaluation, Office of Instructional and Management Services, at the University of Illinois at Urbana-Champaign. He holds an M.A. in educational psychology from Michigan State University and a Ph.D. in educational measurement from the University of Iowa.

He has written numerous articles and book chapters in the area of instructional evaluation, including computer-related educational applications. He is also a co-author of *Evaluation of Corporate Training Programs.* Brandenburg serves on the technology council of the National Society for Performance and Instruction and is a consulting editor for its refereed journal, *Performance Improvement Quarterly.*

C. Victor Bunderson
Vice-President for Research Management
Educational Testing Service
Princeton, NJ 08540
609/734-1806

Victor Bunderson is currently vice-president for research management at Educational Testing Service in Princeton. His work has focused on developing and introducing significantly improved and widely applicable education and training systems using advanced hardware, software, and human technologies.

Bunderson's work began in academic settings, including the University of Texas and Brigham Young University. He also played a major role in the start-up of several companies involving advanced computer and information technologies and education and training applications, including the WICAT Education Institute, WICAT Systems and Knowledge Engineering, Inc. Bunderson holds a doctorate in psychology from Princeton University.



Gebreyesus Hamda
Senior Human Resource Development Analyst
Florida Department of Transportation
Office of Human Resource Development
605 Suwannee St., M.S. 56
Tallahassee, FL 32399-0450
904/487-3917

"Gabe" Hamda is a senior human resource development analyst for the Florida Department of Transportation, Office of Human Resource Development. He is responsible for needs assessments, program evaluation, employee career development, research and development, and strategic planning. Previously, Hamda was on the research staff at the Center for Needs Assessment and Planning at Florida State University in Tallahassee. He is also a former faculty lecturer at the Bahr Dar Teachers' College, Addis Ababa University in Ethiopia.

Hamda holds a B.Ed. in pedagogical sciences from the Academy of Pedagogy of Ethiopia and an M.S. and a Ph.D. in instructional systems from Florida State University. He also holds a graduate certificate in human resource development from Florida State University.

C. W. Hoover Jr.
Executive Director
Interconnections and Power Technology Division
AT&T Bell Laboratories
260 Cherry Hill Road Parisippany, NJ 07054
718/260-3216

Charles Hoover Jr. is executive director of the Interconnections and Power Technology Division at AT&T Bell Laboratories. He is the former chair of AT&T's Committee on Education, which sets policy and oversees education programs for about 30,000 engineers in the R&D community.

He received a bachelor's degree in mechanical engineering from Yale University in 1946, a B.S. in electronics engineering from the Massachusetts Institute of Technology in 1947, and M.S. and Ph.D. degrees in physics from Yale in 1951 and 1954, respectively.

Hoover has been granted several patents for his inventions and is the author of a number of technical articles. He is a fellow of the Institute of Electrical and Electronics Engineers and a member of the American Physical Society, Tau Beta Pi, and Sigma Xi. He is president and trustee of the Summit Civic Foundation, a non-profit corporation that builds affordable housing in Summit, NJ.

Randolph I. James
Consultant
James & Associates
P.O. Box 1408
Orangevale, CA 95662
916/989-5306

Randolph James, a management consultant, has worked in the field of performance improvement for more than nine years. He holds a graduate degree in psychology (industrial/organizational) from California State University-Sacramento, where he has been a part-time faculty member, teaching Psychology in Personnel Training and Development.

James has published several articles on improving job performance and has presented papers at national conferences and workshops. He is a member of the National Society for Performance and Instruction.

Joseph T. Martelli
Research Scientist
Center for Social and Economic Issues
Industrial Technology Institute
P.O. Box 1485
Ann Arbor, MI
48106 313/769-4000

Joseph Martelli has had significant experience in human resource development, technical training, and organization development, working and consulting in several Fortune 500 organizations. Prior to joining the Industrial Technology Institute, Martelli was the top technical training official for Kellogg in Battle Creek, Michigan. At Kellogg, he was responsible for planning and executing the technical training systems for a new $200 million cereal manufacturing facility. He also has worked on plant equipment start-ups in the pulp and paper, petroleum, and steel industries.

Martelli taught personnel management at Kellogg Community College. He regularly publishes in various training and development periodicals and has made presentations at several national conferences.

He received a B.S. in education and psychology from Kent State University in 1975, an M.Ed. in career and technology education from Bowling Green State University in 1977, and a doctorate in industrial technology from the University of Northern Iowa in 1981.

Sandra J. Merwin
President
Sandra Merwin Enterprises
4655 Baker Road
Minnetonka, MN 55345
612/935/1446

Sandra Merwin, president of Sandra Merwin Enterprises, designs and develops training programs for business and education. For the past 10 years she has been involved in research concerning training and evaluation. Her book, *Effective Evaluation Strategies and Techniques*, addresses an integrated evaluation system for classroom training.

In 1985, Merwin received the Professional Excellence Award for Technological Innovation from the Southern Minnesota Chapter of the American Society for Training and Development for her research efforts.

Nancy R. Mosier
Manager, Education Consulting
Arthur Andersen & Co.
1666 K Street N.W.
Washington, DC 20006
612/334-3636

Nancy Mosier is a manager in the education consulting practice of Arthur Andersen & Co. She has worked with clients in a variety of industries and has managed projects in all phases of instructional system design. Prior to joining Arthur Andersen & Co., Mosier was a product manager for Control Data Corp., where she designed and managed computer-based training courses in health awareness and academic education.

Mosier holds B.S. and M.A. degrees in journalism and is a Ph.D. candidate in employee training and organization development at the University of Minnesota.

Brian P. Murphy
President
The HRD Department, Inc.
2187 Dayton Avenue
St. Paul, MN 55104
612/642-1427

Brian Murphy co-owns and operates The HRD Department, Inc., a company that leases human resource development services and programs to business and industry. He has held positions in the private sector in human resource planning, training and development, marketing education, and technical training. He also has developed academic curricula for use in vocational and public schools and has instructed at the university level. Prior to his professional pursuits, Murphy worked in the skilled trades as an electronics technician, electrician, machinist, and exhibits technician.

Murphy earned a B.S. in industrial education and an M.Ed. in career and technology education from Bowling Green State University. He has been active professionally in both the training and education fields as an author and officer in professional societies. He has written articles for professional journals, including the *NAITTE Journal*, *Performance and Instruction*, and *Training & Development Journal*. He has given numerous presentations to groups such as the American Vocational Association, the National Society for Performance and Instruction, the American Society for Training and Development, and the Minnesota State Department of Education.

Jeffrey D. Oberlin
Manager of Planning and Evaluation
Motorola Training and Education Center
Motorola, Inc.
1303 E. Algonquin Road
Schaumburg, IL 60196
312/576-2899

Jeff Oberlin is manager of planning and evaluation at the Motorola Training and Education Center. He has served as program development manager, instructional video producer, and design manager at AT&T and Motorola. Among his most interesting assignments were conducting a benchmark study of employee development systems in major Japanese electronic firms and directing an Asia-Pacific strategic planning conference in Penang, Malaysia.

Oberlin earned a bachelor of arts degree in telecommunications and a master's degree in instructional systems technology from Indiana University in Bloomington. He has also studied behavioral analysis in training at the Huthwaite Research Group in Sheffield, England.

Gene W. Poor
Vice-President, Good Display, Inc.
Associate Professor
College of Technology
Bowling Green State University
Bowling Green, Ohio 43404
419/372-2437

Gene Poor is vice-president of Good Display, Inc. in Toledo, Ohio and associate professor of visual communications at Bowling Green State University. He has written publications in the areas of training, organizational communication, and creativity. His most recent research and development has focused on computer-driven, lifelike robots for the training and exhibit industry.

Poor received B.S. and M.Ed. degrees in industrial education from Kent State University. He earned his Ph.D. in educational administration from Bowling Green State University.

Suzanne Saxe
Manager, Management and
Employee Development
Central Training Group
Wells Fargo Bank & Co.
79 New Montgomery Street, 5th Floor
San Francisco, CA 94115
415/383-3955

Suzanne Saxe is manager of the Management and Employee Development Central Training Group at Wells Fargo Bank & Co. She is responsible for design, development, and delivery of management and supervisory training and a variety of other training programs.

In addition, Saxe has taught at major universities in the San Francisco Bay Area and has published articles on instructional design and adult education. In 1986, she received the Dissertation of the Year award from the American Society for Training and Development. Saxe holds an Ed.D. in curriculum and instruction from the University of San Francisco.

Scott A. Sayre
Director
Sayre Media Consultants
89 N. Victoria Sreet
St. Paul, MN 55104
612/290-0042

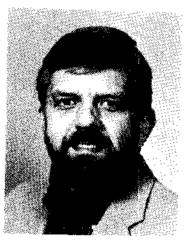
Scott Sayre is a media production and applications consultant in education and industry. His clients include the Technology for Literacy Project, Honeywell, and Ball Corporation. Sayre's current projects include developing a computer-based system for designing and pilot testing interactive video programs and assisting in the development of an interactive videodisc program to assist trainers and educators in selecting appropriate distance delivery systems.

Currently Sayre is completing his doctorate in industrial education at the University of Minnesota, specializing in training and development. He received his M.Ed. degree in career and technology education and a B.A. in visual communications technology from Bowling Green State University in Ohio.

Catherine M. Sleezer
Training and Development Consultant
5257 Michaele Lane
Minnetonka, MN 55345
612/470-0324

Catherine Sleezer's consulting experience includes economic evaluation of investments in human resource development, managing the HRD function, and assessing corporate culture. She works with a variety of manufacturing and service industries.

Sleezer is a Ph.D. candidate in training and development at the University of Minnesota and is a research assistant for the University's Training and Development Research Center. She has published articles in a number of training and development journals.



Richard A. Swanson
Professor and Director
of the Training and Development Research Center
University of Minnesota
1954 Buford Avenue
St. Paul, MN 55108
612/624-9727

Richard Swanson is a professor and the director of the Training and Development Research Center at the University of Minnesota. He also serves as a member of the American Society for Training and Development Research Committee and is the editor of the newly formed research journal, *Human Resource Development Quarterly.* (HRDQ), which is sponsored by ASTD and Jossey-Bass, Publishers, Inc.

Swanson is the co-author of *Forecasting Financial Benefits of HRD* and *Performance at Work* and is also the past editor of *Performance and Instruction Journal* and the *Journal of Industrial Teacher Education.* He received his Ed.D. in industrial education from the University of Illinois.

Michael T. Wood
Manager
Training and Technical Assistance Group
Industrial Technology Institute
2901 Hubbard Road
Ann Arbor, MI 48106
313/769-4563

Michael Wood is manager of the Training and Technical Assistance Group at the Industrial Technology Institute (ITI). Prior to joining ITI, he directed the Center for the Study of Social Change at the Human Affairs Research Centers of the Battelle Memorial Institute. He has also served on the psychology and business faculties of Ohio State University, the University of Washington, and Pacific-Lutheran University.

Wood is a past recipient of the American Psychological Association Dissertation Award and has written articles on issues of organizational behavior, leadership, and training and development. He received his Ph.D. in industrial/organizational psychology from the University of Illinois.