Levine, E. L. (1983). *Everything you always wanted to know about job analysis.* Tampa, FL: Mariner Publishing Co.

Maye, D. M., & Goldstein, I. L. (1989, June). *Content validity with the consent of justice: A collaborative methodology.* Paper presented at the International Personnel Management Association Conference, Orlando, FL.

McCormick, E. J. (1979). *Job analysis: Methods and applications.* New York: AMACOM.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *American Psychologist, 89,* 575–588.

Prien, E. P. (1977). The function of job analysis in content validation. *Personnel Psychology, 30,* 167–174.

Prien, E. P., Goldstein, I. L., & Macey, W. H. (1987, August). Multi-domain job analysis: Procedures and applications in human resource management and development. *Training and Development Journal,* pp. 66–72.

Sackett, P. R. (1987). Assessment centers and content validity: Some neglected issues. *Personnel Psychology, 40,* 13–25.

Schmitt, N. (1987, April). *Areas of Continued Debate in Personnel Selection: Principles III.* Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Schmitt, N., & Cohen, S. A. (1989). Internal analyses of task ratings by job incumbents. *Journal of Applied Psychology, 74,* 96–104.

Schmitt, N., & Landy, F. (1992). Internal analyses of task ratings by job incumbents. In N. Schmitt & W. Borman (Eds.), *Personnel selection.* San Francisco: Jossey-Bass.

Schmitt, N., & Noe, R. A. (1983). Demonstration of content validity: Assessment center example. *Journal of Assessment Center Technology, 6,* 5–11.

Schneider, B., & Konz, A. M. (1989). Strategic job analysis. *Human Resource Management, 28,* 51–63.

Society for Industrial and Organizational Psychology. (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.). College Park, MD: Author.

Tenopyr, M. L. (1977). Content-construct confusion. *Personnel Psychology, 30,* 47–54.

2

A Theory of Performance

John P. Campbell, Rodney A. McCloy,
Scott H. Oppler, Christopher E. Sager

In Schmitt & Borman (1993)

In this chapter we examine alternative models for the substantive content and latent structure of job performance as a construct and also consider their critical measurement and research implications. The discussion is based on two premises: first, that individual performance on a "task," virtually any task that the culture views as having value, is one of the most important dependent variables in psychology, basic or applied; and second, that the word *performance* is misused and exploited to the extreme in society at large, and is frequently butchered beyond recognition in psychology. If the dependent variable is the variable of *real* interest and if performance is perhaps our most important dependent variable, then more often than not we simply may not know what our real interest is. If we want to accumulate knowledge about how to measure, predict, explain, and change performance but have no common understanding of what it is, then building a cumulative research record is difficult to impossible and industrial and organizational (I/O) psychology is in for continued unfavorable comparisons with other sciences.

Even a brief examination of applied research illustrates the diversity of measurement operations that have been labeled, explicitly or implicitly, as measures of performance. Consider just the following:

- Time to complete a training course
- Grades or achievement test scores earned in training
- Number of errors made in a simulator
- Number of Tinkertoy figures assembled in a forty-five-minute experimental session
- Number of one-minute marketing interviews completed outside a shopping center in one day
- Number of pieces produced
- Number of defective pieces produced
- The total or average cost of the pieces produced
- Number of proposals written
- Total value of contracts won
- Total value of sales
- Number of grievances or complaints incurred
- Length of tenure in the organization
- Total days absent
- Salary level
- Promotion rate within an organization
- Percentage over budget
- Supervisor, peer, subordinate, or self ratings of "overall" performance
- Scores on a paper-and-pencil job knowledge test
- Scores on a professional certification test
- Number of citations in the citation index over a three-year period
- Number of refereed journal articles published in a six-year period

Even this brief list encompasses a wide variety of phenomena under one label. For any researcher faced with such helter-skelter in any domain, a reasonable response would be to reach for the available theory for guidance as to which things are important and which are not, how the important variables should be defined and what they are good for, and what the general operations of measurement may be. For example, if our concern is with the nature of cognitive abilities (Horn, 1989), personality (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990),

motivation (Kanfer, 1990), or even leadership (Bass, 1990), there are literally decades of research and theory to consult. With regard to performance itself, however, there is virtually none. Performance as a construct has received very little research or theoretical attention.

Two possible reasons for the deficit come to mind. The first is that it always seems to be the independent variable that generates the most professional or scientific interest. People can get goose bumps over a new treatment (such as empowered work groups) or a new ability variable (for example, tacit knowledge), but spending a lot of time or resources to understand performance itself seems to be not very exciting or fundable. A second possibility is that most of us hold a strong stereotypical assumption that the definition of performance and the designation of its indicators are things that get taken care of by somebody else, that they are really out of our hands.

We think the latter point of view supports the one model of performance that has dominated applied research during most of the last eighty to ninety years, at least in the world of work. Be it explicit or implicit, a theory is always there (Campbell, 1990b).

The Classic Model as Villain

The classic theory, or default position, that has dominated applied research for much of the last century says simply that performance is one thing; that is, the general factor will account for almost all the relevant true-score covariances among observed measures. Covariance matrices that appear otherwise do so because of differential reliabilities, the influence of method-specific variance, or other kinds of contamination. Thus the goal of measurement is to obtain the best possible measure of the general factor. For job performance, the classic view specifies that the best possible measure is an "objective" indicator of individual accomplishment that is maintained by the organization itself, such as number of pieces produced or dollar volume of sales.

The reasons for the above characterization are as follows. First, for most of this century the single-criterion measure has

dominated in personnel research. Further, in the scientific and professional literature the term *job performance* is virtually always used in the singular, with no explicit or implicit conditionals. When we discuss whether or not job performance can be changed by this or that treatment, predicted by a particular ability, or measured better or worse by a particular method, the implication is clearly that there is one general thing to be changed, predicted, or measured. This certainly seemed to be what Thorndike (1949) had in mind when he coined "immediate," "intermediate," and "ultimate" criteria relative to the validation of aptitude tests in occupational or educational settings. This was a very unfortunate and counterproductive characterization of the goals of performance measurement, but it is incorporated in virtually every I/O psychology textbook written between 1950 and 1990.

This point of view is also the major source of the hated "criterion problem," which laments that if the best way to measure individual job performance is with objective indicators of individual accomplishments that organizations keep in their archives, then we indeed have a problem, because the search for reliable, uncontaminated, objective indicators that significantly reflect the ultimate criterion has been a failure. Good ones are never found, and this holds up progress. If only we had a good criterion, *then* you would really see something. The fallback position that is forced on us is to use supervisory ratings, which are "subjective," contaminated with halo error, and full of information processing errors. Consequently, the field must bear a major burden of guilt for being unable to find, or appropriately transform, the objective archival record that surely must be there.

The gist of this chapter is that (1) the general factor cannot possibly represent the best fit, (2) the notion of an ultimate criterion has no meaning, (3) the subjective versus objective distinction is a false issue, and (4) there is a critically important distinction to be made between performance and the results of performance.

We are certainly not the only ones to complain about the classic model creating a lot of trouble (for example, Dunnette,

1963; Wallace, 1965). Upwards of fifty years ago, there were numerous calls to spend as much time and energy developing, studying, and theorizing about criteria as were spent on the predictor (Jenkins, 1946; Nagle, 1953; Patterson, 1946; Toops, 1944). Toops (1944) typified this sentiment: "The criterion . . . seldom receives half the time or attention it requires or deserves. If the criterion is slighted, the time spent on the tests is, by so much, largely wasted" (p. 290). Often, criteria were taken as given, an action that "overcame the problem of criterion development by ignoring it" (Nagle, 1953, p. 271). Wallace (1965) urged that construct validation be used to understand criteria better, and James (1973) described how construct validation could be applied to criteria as easily as to predictors.

However, such pronouncements by a few respected figures seem to have had very little impact, so why go through the same exercise again? The principal reason is that we think the cumulative record now makes it possible to offer much more substantive alternatives to the one-factor model. Also, recent developments in restricted factor models, confirmatory analysis, and the like provide a better language for talking about the latent structure of the performance domain and provide a clearer way of analyzing relevant data. It is our firm conviction that unless I/O psychology begins to argue energetically about the substantive characteristics of performance, the most critical dependent variable in the entire field will forever remain unspecified. Consequently, we propose a substantive model that specifies the content of performance, its direct determinants, and its critical dynamic properties. The implications of the model for measurement and research are also discussed. The description of the model itself is an expanded version of that presented by Campbell (1990a).

A Proposed Theory of Performance

Although our focus is on job performance, we think the basic implications are much broader and illustrate this point as we go along.

Performance Defined

A full understanding of job performance depends on having some understanding of the organizational goals to which the individual is supposed to contribute. For example, the issues surrounding the instruction versus research goal in universities are familiar to most of us. If this sounds too ho-hum, then get a bit closer to life and death and consider, for air traffic controllers, the two goals of maximizing air traffic safety and maximizing efficiency of air traffic movement *at the same time.* Even closer to home, what are, or should be, the goals of one's family? Family members are always judging each other in terms of their contribution to the family's goals. Some families consider them explicitly, others do not; but goals are always there, if only by default. In the job setting, both confusion and trouble can result if either the formal employment contract or the informal psychological contract incorporates goals different from those against which performance is actually judged (Rousseau, 1989; Schein, 1970).

Performance is herein defined as synonymous with behavior. It is something that people actually do and can be observed. By definition, it includes only those actions or behaviors that are relevant to the organization's goals and that can be scaled (measured) in terms of each individual's proficiency (that is, level of contribution). Performance is what the organization hires one to do, and do well. Performance is *not* the consequence or result of action, it is the action itself. Admittedly, this distinction is troublesome in at least one major respect—behavior is not always observable (for example, cognitive behavior, as in solving a math problem) and can be known only by its effects (for instance, producing a solution after much "thought"). However, "solutions," "statements," or "answers" produced as a result of covert cognitive behavior and totally under the control of the individual are included as actions that can be defined as performance. In general, a strict definition of "observable behavior" is epistemologically difficult. Let us simply argue that performance consists of goal-relevant actions that are under the control of the individual, regardless of whether they are cognitive,

motor, psychomotor, or interpersonal. Consequently, writing a job-relevant memo falls within the definition, if the availability of a word processor is a constant, whereas the number of pieces produced does not, unless such an outcome is under the complete control of the individual.

For the model espoused here, it is axiomatic that job performance is not just one thing. A job, *any* job, is a very complex activity; and, for *any* job there are a number of major performance components that are distinguishable in terms of their determinants and covariation patterns with other variables. Some examples of performance components are giving emergency first aid (police officer), planning and designing undergraduate courses (university faculty), driving safely under hazardous conditions (truck operator), rewarding sales personnel for appropriate actions (sales supervisor), and using rules of separation efficiently (air traffic controller).

Performance is to be distinguished from *effectiveness* and from *productivity. Effectiveness* refers to the evaluation of the results of performance. By definition, the variance in a measure of effectiveness is controlled by more than the actions of the individual. Dollar amount of sales is an obvious example. An implication of this distinction is that rewarding or punishing individuals on the basis of effectiveness may be unfair and counterproductive. As an indicator of performance, effectiveness is by definition contaminated. This is not to argue that results are not important. They most certainly are. Ultimately, the organization needs to know the sources of variation in performance *and* the sources of variation in effectiveness. Effectiveness is the bottom line, and organizations cannot exist without it. By defining performance this way we are simply arguing the point that if the research questions deal with predictor validities, or training effects, or any other strategy focused on the individual, then the dependent variable should not be something that the individual cannot influence. Further, it would be maximally informative to know the relationship of performance to effectiveness, and not to confound them.

The usually agreed upon definition of productivity (Mahoney, 1988) is the ratio of effectiveness (output) to the cost of

achieving that level of effectiveness (input). Its primary use is as a relative index of how well a group, organization, or industry is functioning. Depending on which inputs (costs) are used in the denominator, it is possible to talk about the productivity of capital, the productivity of technology, or the productivity of labor. That is, total productivity has its subcomponents.

For the sake of completeness, *utility* is defined as the value of a particular level of performance, effectiveness, or productivity. That is, we can talk about the utility of performance, the utility of effectiveness, or the utility of productivity. For example, for a board of directors the utility of particular levels of profit (an effectiveness measure by our definition) may bear no resemblance to its dollar amount. At least under the federal tax laws existing in the late 1980s, a high level of profit may have negative utility because it invites takeover attempts and ultimately leads to a heavy burden of corporate debt. In the school system, what is the relative value of high student proficiency in algebra versus high proficiency in English? For individual job holders, what is the value of a top performer versus an average one? All of these "utilities" may or may not be expressed in dollar terms. In certain kinds of organizations it may make little sense to use a dollar metric (Sadacca, Campbell, DiFazio, Schultz, & White, 1990). Also, the regression of utility on performance or effectiveness may not be linear (Pritchard & Roth, 1991).

Determinants of Individual Differences in Performance

We next must distinguish among performance *components* (performance *factor* and performance *construct* are regarded as synonymous terms), performance *determinants,* and the *antecedents* or *predictors* of performance determinants. After making these general distinctions, we propose a substantive taxonomy of performance components. The variables and relationships to be examined are summarized in Table 2.1.

The performance components, or factors, are the distinguishable categories of things people are expected to do in a job. They are the latent variables. Although their true-score inter-

Table 2.1. Determinants of Job Performance Components.

$PC_i = f$ [*Declarative Knowledge (DK)*	×	*Procedural Knowledge and Skill (PKS)*	×	*Motivation (M)*]
Facts		Cognitive skill		Choice to perform
Principles		Psychomotor skill		Level of effort
Goals		Physical skill		Persistence of effort
Self-knowledge		Self-management skill		
		Interpersonal skill		

$i = 1, 2, \ldots, k$ performance components

Predictors of Performance Determinants[b]

$DK = f$[(ability, personality, interests), (education, training, experience), (aptitude/treatment interactions)]

$PKS = f$[(ability, personality, interests), (education, training, practice, experience), (aptitude/treatment interactions)]

$M = f$(whatever independent variables are stipulated by your favorite motivation theory)

Note: This entire schema can be repeated for educational performance, training performance, and laboratory task performance.

[a]Obviously, performance differences can also be produced by situational effects such as the quality of equipment, degree of staff support, or nature of the working conditions. For purposes of this model of performance, these conditionals are assumed to be held constant (experimentally, statistically, or judgmentally).

[b]Individual differences, learning, and motivational manipulations can only influence performance by increasing declarative knowledge or procedural skill or by influencing the three choices (that is, influencing motivation).

correlations are not zero, they are low enough to yield a significantly different rank ordering of people depending on the performance component being talked about.

Individual differences on *each* specific performance component are viewed as a function of three, and only three, major determinants—declarative knowledge, procedural knowledge and skill, and motivation. Declarative knowledge is simply knowledge about facts and things. Specifically, it represents an understanding of a given task's requirements, such as general principles for equipment operation (Anderson, 1985; Kanfer

& Ackerman, 1989). Procedural knowledge and skill are attained when declarative knowledge (knowing what to do) has been successfully combined with knowing how to do it (modified from Anderson, 1985, and Kanfer & Ackerman, 1989).

As a direct determinant of performance, *motivation* is herein defined as a combined effect from three choice behaviors: (1) choice to expend effort, (2) choice of level of effort to expend, and (3) choice to persist in the expenditure of that level of effort. These are the traditional representations for the direction, amplitude, and duration of volitional behavior. The important point is that the most meaningful way to talk about motivation as a direct determinant of behavior is as one or more of these three choices.

Antecedents of the Performance Determinants

Accounting for individual differences in knowledge, skill, and choice behavior encompasses a very large number of research topics that are not to be discussed here. From the trait perspective, almost a century of research has produced taxonomic models of abilities, personality, interests, and personal histories. Another major research tradition has focused on instructional treatment. At least three major types of such treatments are relevant in the job performance content—formal education, job-relevant training (formal and informal), and previous experience. The possible antecedents of motivation, or choice behavior, are specified by the various theories of motivation. For example, an operant model stipulates that the reinforcement contingency is the most important determinant of the choices people make. Cognitive expectancy models say that certain specific thoughts (for example, self-efficacy, instrumentality, valence) govern these three choices. Other models see such choices as a function of certain stable predispositions, such as the need for achievement. For example, perhaps certain kinds of people virtually always go to work on time and always work hard.

A few general points are important. First, the precise functional form of the $PC = f(DK, PKS, M)$ equation (Table 2.1) is obviously not known and perhaps not even knowable. Further,

spending years of research looking for it would probably not be of much use. Instead, consider the following. Performance will not occur unless there is a choice to perform at some level of effort for some specified time. Consequently, motivation is *always* a determinant of performance, and a relevant question for virtually any personnel selection problem is how much of the variance in choice behavior can be accounted for by stable predispositions measurable at the time of hire and how much is a function of the motivating properties of the situation or the interaction. Performance that is not simply trial and error also cannot occur unless there is some threshold level of procedural skill, and there may indeed be a very complex interaction between *PKS* and *M*. For example, the higher the skill level, the greater the tendency to choose to perform, but skill level may have no relationship with the choice of effort level. That is, the three choices may be controlled by different antecedents.

Another reasonable assumption is that declarative knowledge is a prerequisite for procedural skill (Anderson, 1985). That is, before being able to use the procedural skills that are necessary for task performance, one must know what should be done. However, this point is not without controversy (Nissen & Bullmer, 1987), and it may indeed be possible to master a skill without first acquiring the requisite declarative knowledge. Two examples that come to mind are modeling the social skills of your parents and modeling the "final form" of an expert skier without really "knowing" what you are trying to do. Nevertheless, given the current findings in cognitive research, the distinction is a meaningful one. Performance could suffer because procedural skill was never developed or because declarative knowledge was never acquired or because one or the other has decayed. Also, some data suggest that the abilities that account for individual differences in declarative knowledge are different from those that account for individual differences in procedural skills (Ackerman, 1988). At this point, the major implication is still that performance is directly determined only by some combination of these three elements.

Again, the functional relationships between individual differences on the three determinants and individual differences

on a component of performance are not constrained to any particular form. That is, they could be linear or nonlinear, and even exhibit strong asymptotes. The functional relationships for declarative knowledge and procedural skill are most likely monotonically increasing, although the same may not be true for the choice determinants.

A Taxonomy of Major Performance Components

If performance is more than one thing, then what are the major parts? Across the entire occupational spectrum how many parts are there? Thousands? Hundreds? A few dozen? Is the latent structure hierarchical, such that there are a small number of general factors at the highest level and increasing specificity as one goes down the hierarchy? What is the variation in the form of the latent structure across jobs? Can a "population" of jobs be defined by a common performance structure? Is there a general performance factor? For some jobs but not for others? A sampling of such questions raises the problem of whether the inherent complexity of performance makes investigating its latent structure counterproductive. The assumption here is no.

The model described is a factor model, but it invokes no general factor. Instead, a hierarchy is proposed that has eight factors at the most general level. The eight factors are intended to be sufficient to describe the top of the latent hierarchy in all jobs in the *Dictionary of Occupational Titles*. However, the eight factors are not of the same form. They have different patterns of subsidiary general factors, and their content varies differentially across jobs. The critical constant is that the manifest representation of each factor must be in terms of things people do, not the "bottom line" or the "results" of what they do. A brief explication of each of the eight factors is given below.

1. *Job-specific task proficiency.* The first factor reflects the degree to which the individual can perform the core substantive or technical tasks that are central to the job. They are the job-specific performance behaviors that distinguish the substantive content of one job from another. Constructing custom kitchens, doing word processing, designing computer architecture, driving a city bus through Chicago traffic, and directing air traffic are all categories of job-specific task content. Individual differences in how well such tasks are executed are the focus of this performance component.

2. *Non-job-specific task proficiency.* This factor reflects the situation that in virtually every organization, but perhaps not all, individuals are required to perform tasks or execute performance behaviors that are not specific to their particular job. For example, in research universities with Ph.D. programs, the faculty must "teach classes," "advise students," "make admission decisions," and "serve on committees." All faculty must do these things, in addition to doing chemistry, psychology, economics, or electrical engineering. In the military services this factor is institutionalized as a set of common tasks (for example, first aid, basic navigation, using NBC equipment) for which everyone is responsible.

3. *Written and oral communication task proficiency.* Many jobs in the workforce require the individual to make formal oral or written presentations to audiences that may vary from one to tens of thousands. For those jobs the proficiency with which one can write or speak, independent of the correctness of the subject matter, is a critical component of performance.

4. *Demonstrating effort.* The fourth factor is meant to be a direct reflection of the consistency of an individual's effort day by day, the frequency with which people will expend extra effort when required, and the willingness to keep working under adverse conditions. It is a reflection of the degree to which individuals commit themselves to all job tasks, work at a high level of intensity, and keep working when it is cold, wet, or late.

5. *Maintaining personal discipline.* The fifth component is characterized by the degree to which negative behavior such as alcohol and substance abuse at work, law or rules infractions, and excessive absenteeism are avoided.

6. *Facilitating peer and team performance.* Factor six is defined as the degree to which the individual supports his or her peers, helps them with job problems, and acts as a de facto trainer.

It also encompasses how well an individual facilitates group functioning by being a good model, keeping the group goal-directed, and reinforcing participation by the other group members. Obviously, if the individual works alone, this component will have little importance. However, in many jobs, high performance on this factor would be a major contribution toward the goals of the organization.

7. *Supervision/leadership.* Proficiency in the supervisory component includes all the behaviors directed at influencing the performance of subordinates through face-to-face interpersonal interaction and influence. Supervisors set goals for subordinates, they teach them more effective methods, they model the appropriate behaviors, and they reward or punish in appropriate ways. The distinction between this factor and the previous one is a distinction between peer leadership and supervisory leadership. Although modeling, goal setting, coaching, and providing reinforcement are elements in both factors, the belief here is that peer versus supervisor leadership implies significantly different determinants.

8. *Management/administration.* The eighth and last factor is intended to include the major elements in management that are distinct from direct supervision. It includes the performance behaviors directed at articulating goals for the unit or enterprise, organizing people and resources to work on them, monitoring progress, helping to solve problems or overcome crises that stand in the way of goal accomplishment, controlling expenditures, obtaining additional resources, and representing the unit in dealings with other units.

These eight factors are meant to be the highest-order factors that can be useful. To reduce them further would cover up too much. However, it is also acknowledged that not all the factors are relevant for all jobs. For example, not all jobs have a supervisory or management component. Not all jobs involve formal communication tasks. Some job holders have no peers or work group. What the model asserts is that the eight components, or some subset of them, can describe the highest-order latent variables for every job in the occupational domain. Further, three of the factors—core task proficiency, demonstrated effort, and maintenance of personal discipline—are major performance components of *every* job.

Direct evidence for this taxonomy is admittedly sparse. We have been most influenced by the long-term Selection and Classification Project (Project A) sponsored by the U.S. Army (Campbell & Zook, 1990), which settled on a five-component model of performance (Campbell, McHenry, & Wise, 1990) for a population of 275 entry-level skilled jobs. The model received confirmation in a follow-up of this same sample three years later, after reenlistment (Campbell, 1991), and will be tested again on another large sample of entry-level personnel from a different cohort.

Supporting evidence comes from studies of performance using the Behaviorally Anchored Rating Scales (BARS) technique (for example, Campbell, Dunnette, Arvey, & Hellervik, 1973), which generally seem to produce factors that can be fit into the taxonomy, and from a recent study by Hedberg (1989) that used extensive interview and critical incident methodology to describe various parameters of feedback as it occurs in an ongoing job environment. A content analysis of performance feedback episodes for a sample of financial analysts and project managers tended to mirror the eight-factor solution quite closely. Also, a great deal of evidence for some of the factors comes from descriptive research on what supervisors, leaders, and managers do (for example, Yukl, 1987; Bass, 1990).

In general, however, research on the criterion problem has tended to avoid any formal investigation of the latent structure of performance. A type of study that is *not* appropriate for this purpose is a factor analysis of some mixture of performance measures, indices of effectiveness, and perhaps even measures of performance determinants in the same matrix.

If the eight-factor representation is an appropriate one, then it should receive support from future construct validation efforts, as against alternatives; and it should show consistent differential correlations with other variables. For example, the prediction equations for each factor should be recognizably different.

Implications of the Eight-Factor Latent Structure

The nature of the lower-order factors within each of the eight major components has been the subject of considerable research for some of them and is a matter of speculation for others. If we generalized from the classic situation-specificity hypothesis in personnel psychology, then we might expect the number of subfactors for component 1 (job-specific task proficiency) to equal the number of jobs. That is, the prediction equation for component 1 would be different for each job. Obviously, we already know from lots of data that this would be a very poor description of the latent structure. However, virtually no one would expect the prediction equation for component 1 to be the same for jazz musicians, graphic artists, PGA tour golfers, NFL defensive linemen, theoretical economists, Protestant clergy, farm managers, long-haul truck operators, stockbrokers, and air traffic controllers. Where between these two extremes is a more appropriate description of the latent structure of factor 1? The model proposed here assumes that the number of discriminable subfactors for component 1 is a manageable number, and that it would be quite possible to build up a systematic body of knowledge around the major differences in the correlates for the subfactors.

Also, it is interesting to speculate about the subfactors that might be found in factor 4. In particular, should there be a distinction between physical effort and cognitive effort? A possible confusion here is the distinction between effort behaviors as a component of performance and the three choices as a determinant of performance. As a component of performance, demonstrated effort consists of a set of behaviors that represent mental or physical effort expenditure that contributes to the organization's goals. As with any other performance component, demonstrated effort is a function of declarative knowledge, procedural knowledge and skill, and motivation. For example, someone may not have worked overtime because they did not "know it was an option."

To the limited extent possible, the eight basic components have been defined so as to remove causal relations among them.

However, the true structure is probably not so static. For example, it is most likely the case that being very accomplished on the core tasks of the job or being a very proficient writer or speaker enhances peer or supervisory leadership performance. This is not the same as saying that task performance and supervisory performance have common determinants (such as certain procedural skills). Instead, the supervisor's manifest task performance itself may change the perceptions of the subordinate and thereby alter the influence process.

Finally, we emphasize again that performance and the direct determinants of performance are independent constructs. Because we are just in the beginning stages, future work will most certainly show that this initial taxonomy of performance components is not the "best" description of the latent structure. However, that by itself will not change our view of how to think about the three determinants, which have a much richer history in psychological theory and research.

Critical Performance Parameters

The experimental literature on task performance identifies several performance parameters that may have particular relevance for this kind of model.

Speed Versus Accuracy. One such parameter is speed versus accuracy or quantity versus quality. The speed-accuracy distinction is important because it has been shown that speed and level scores for the same task load on independent factors (for example, Kyllonen, 1985) and speed of problem solving may not correlate with accuracy of problem solving in the same content domain (Lohman, 1989). As a consequence, it seems reasonable to expect that predicting quantity (speed) and predicting quality (accuracy) would require different equations and that more specific abilities may be more appropriate for the former than the latter. The issue in the job performance context encompasses two major questions. First, what kind of trade-off between the two does the organization value? Second, what trade-off is the individual actually making?

Automatic Versus Controlled Processing and the Performance Asymptote. A second parameter of interest is a recognizable phenomenon that can occur during cognitive, psychomotor, or motor skill acquisition and is now referred to as the development of "automaticity" regarding skills or components of skills (Shiffrin & Schneider, 1977). Automaticity means that a skill or a skill component can be performed automatically without investing very much, if any, conscious attentional effort. The individual literally does not have to "think about" performing the task. Most of us experience such a shift to automaticity when learning to read, ride a bicycle, or type. A controlled process is task performance that requires a significant investment of continuous attention. Performance moves from controlled to automatic processing as a function of learning, experience, and practice, and individuals differ in terms of the speed with which they make the transition. Tasks with the potential for being performed automatically are known as consistent tasks. At least in theory, consistent tasks have a performance asymptote, whereas controlled tasks do not.

Ackerman (1987) and Kanfer and Ackerman (1989) have used these notions to help explain the differential relationship of abilities to technical task performance at different stages of training or experience. For example, Ackerman (1987) has modified the basic Fleishman (1972) hypothesis that specific abilities become more important predictors of performance (relative to general ability) as training progresses by showing this to be true only for consistent tasks. That is, general abilities are what count when a great deal of controlled processing is necessary, but under conditions of automaticity it is the more specific abilities that predict individual differences in task performance. For inconsistent tasks, general abilities will always be major predictors. Similarly, the increase or decrease in performance variance as a function of training is seen as dependent on the degree of task consistency. Consistent tasks should show decreasing variance as a function of additional training or experience, and variability between persons on inconsistent tasks should become greater (Ackerman, 1987).

Peak Versus Typical Performance. Another major parameter is peak versus typical performance. In a very illustrative study using supermarket checkout personnel, Sackett, Zedeck, and Fogli (1988) obtained the correlation between a standardized work sample measure, administered by the researchers, and an on-line computerized record of actual performance on the very same job tasks. Both measures were highly reliable, but the correlation between the two was surprisingly low. The authors called this a distinction between maximum and typical performance and reasoned that the cause of the lowered correlation was the uniformly high motivation generated by the standardized situation versus the differential motivation across individuals in the actual work setting. If such an explanation is accurate, then attempts to model performance must face the issue of what to do with the distinction. The proposed model stipulates that both measures address the content of factor 1 (core technical performance), but the standardized job sample tries to hold the motivation determinants constant for all individuals (at a high level), whereas the archival records do not.

Measurement Considerations

In terms of consistency between the measurement operations and the definition of performance, the proposed model allows only three, or possibly four, primary measurement methods.

Ratings. First, the model strongly implies that ratings (expert judgments) should remain as an important method since they can be constructed to conform directly to the specified definition. The principal worry about ratings has always been that no matter how they are constructed, there may still be significant contamination by systematic variance unrelated to the performance of the person being assessed. Thinking of performance rating as a sequence of observation, sampling, encoding, storage, retrieval, evaluation, differential weighting, and composite scoring allows it to be a very complex cognitive process that allows many opportunities for entry of both unsystematic variance and

contamination (Cooper, 1981a; DeNisi, Cafferty, & Meglino, 1984; Ilgen & Feldman, 1983).

As with any criterion measure, the overall problem is to estimate the proportion of observed variance accounted for by (1) the latent variable, (2) general method variance, (3) measure-specific method variance, (4) other systematic contamination, and (5) unreliability. It has been difficult, however, to assess the effect of nonrelevant sources of variation on the accuracy of job performance ratings. For example, the model clearly implies that it is inappropriate to infer that the variance in ratings that does not overlap with the variance in other measures of performance (say, job samples) is "error variance." As an alternative, some researchers have created videotapes in which the true level of job performance on the dimensions to be rated is scripted (Borman, 1978; Murphy & Balzer, 1989). Accuracy is then assessed by comparing the script to ratings made by raters who view the videotape. However, rating a stranger immediately after viewing a five- to fifteen-minute videotape is a very reduced construction of the situation the rater faces in an organizational setting (for example, Cooper, 1981b; Nathan & Tippins, 1990).

A complete partitioning of the total variance in job performance ratings would require *identifying sources of variance attributable to the rater* that are (1) idiosyncratic (unique to the individual rater), (2) subgroup-specific (unique to raters of the same type), or (3) shared with all raters and *identifying all the significant variables that influence each perspective,* including true score differences on performance. The optimal situation for addressing these requirements would be one in which (1) all raters from all relevant perspectives would rate all incumbents on all dimensions of job performance (Saal, Downey, & Lahey, 1980) and (2) all the variables hypothesized to influence each rater perspective would be assessed. Although this situation is generally not feasible, some partitioning of variance is often possible. For instance, partitioning the contributions due to rater subgroups requires only that each ratee be rated by at least two raters from each subgroup (Sager, 1990). However, partitioning the variance due to idiosyncratic sources requires that each rater also rate the same group of ratees.

Although ratings generally have bad press, the overall picture is not as bleak as might be expected, given all the ways the true-score variance can be contaminated. One advantage of ratings, assuming they share at least some variance with the latent variable, is that their content can be directly linked to the measurement objectives by straightforward content validation methods (for example, critical incident sampling combined with retranslation procedures). Also, *if* used with care, their reliabilities are usually respectable (.50–.60) and can be improved considerably by using more than one rater; and they are as predictable as objective effectiveness measures (Schmitt, Gooding, Noe, & Kirsch, 1984; Nathan & Alexander, 1988). Faith is added by the more thorough attempts to use the method, which have produced credible results (C. Campbell et al., 1990); and over a number of studies, the largest source of variance in ratings is in fact the performance of the ratee (Landy & Farr, 1980). Also, recent studies by McCloy (1990) and Oppler (1990) provide strong evidence that individual differences in performance ratings, when the rating content is carefully defined and raters are trained in their use, are a function of all three performance determinants. The qualifiers are important. Any procedure can be made to look bad by poor implementation, and no method should be made a victim of its most inexpert users. However, it is also true that we still have much to learn about the determinants of ratings.

Standardized Job Samples. The second measurement method allowed by the proposed model is the standardized job sample in which the task content of the job is simulated, or actually sampled intact, and presented to the assessee in a standardized format under standardized conditions. The content validity of the method can also be determined directly, but for the reasons just discussed it may not reflect the influence of all three determinants. That is, individual performance differences on the standardized job sample would be a function of individual differences in declarative knowledge and procedural knowledge and skill, but not motivation. Also, there is always a question about whether the knowledge and skills required by the standardized

sample are different in any major respect from that required in the actual job setting.

Direct Task Observation. The third primary measurement method consists of direct task observation and measurement as it occurs in the job setting. This is what Sackett, Zedeck, and Fogli (1988) were fortunate enough to have for the supermarket checkout personnel. Except in rare instances, this method usually requires rather expensive observational or recording techniques; and for complex work positions the difficulties in observation may be insurmountable.

Alternative Method. In general, a fourth measurement method is dangerous because it equates performance and effectiveness, which is generally forbidden. However, it may sometimes be possible to specify outcomes of performance that are virtually under the complete control of the individual. College grade point average may be such an example; however, saying that could start an argument.

In addition to whether or not the measurement method reflects performance as defined—and not something else—the proposed model identifies two additional critical issues.

First, to avoid criterion content deficiency, the model specifies the population of performance behaviors, along with a procedure for identifying an appropriate sample of content for measurement. Traditionally, this is the domain of job analysis, and a very large literature is available. For example, Project A (C. Campbell et al., 1990) used two principal methods—extensive critical incident descriptions with subsequent retranslation and a series of task analyses that attempted to enumerate every task (that is, the population of tasks) in each job.

The critical incident method is particularly useful because it can sample from the entire population of performance content at the same time that each action is scaled in terms of its contribution to the organization's goals and the major performance components are identified. Also, by changing instructions to the experts providing the incidents, the method can be used to capture performance content for people at various levels of experience, which is a critical parameter. For example, it can be used to contrast "novice" and "expert" performance. Other job analysis methods, such as task analysis, usually cannot provide as much information. Using the critical incident method in this way does not necessarily imply the use of ratings as measurement operations. The critical incident content could be used just as well to specify the content for job observations or simulations.

The second major issue the proposed model addresses is whether the measurement method allows the appropriate determinants (declarative knowledge, procedural knowledge and skill, or motivation) to influence an individual's score. A measurement procedure (for instance, supervisor ratings or scores from a simulator) could choose to control or not control for one or more of them. For example, the measurement objective could stipulate that individual differences in motivation (the three volitional choices) should *not* contribute to individual differences in performance scores, as when evaluating the effects of a skills training program. In such an instance, the measurement goal is to determine whether the specified technical skills have in fact been mastered, not whether the individual chooses to use them in the actual job setting. If the appropriate determinants cannot influence the score on the criterion measure, then the measure is "determinant deficient," which is to be distinguished from the usual use of the term as a label for content deficiency.

Perhaps another example would help illustrate this very fundamental point. It is generally agreed that most commercial airline accidents are the results of faulty "cockpit management." That is, at the critical time there is a breakdown in task delegation, communication, or teamwork. These specific variables seem to represent factors 7 and 8 in the performance component taxonomy. If a simulator is used as a performance measure, then, given that the simulator allows performance on factors 7 and 8 to be observed, a critical question concerns which performance determinants should be allowed to operate and which should be controlled. For example, a critical determinant may be the hesitation of a junior crew member to question the actions of the senior pilot if he or she appears to be in error. To bring this determinant into the simulator, the simulator

declarative knowledge, procedural knowledge and skill, and/or motivation. For example, even though improvements in the quality of supervision can certainly improve individual performance, the mechanisms for the change must be some combination of changes in the individual's declarative knowledge, procedural skill, and choice behavior. The supervisor can "teach" the individual new knowledge or new skills or "motivate" changes in one or more of the three choices. If we do not pay attention to what is happening with regard to the direct determinants, then we may totally misunderstand why better supervision has an effect on individual performance.

3. The proposed model is not a prescription for the best techniques organizations can use to promote higher individual performance. Its intent is to provide a reasonable way of thinking about such issues systematically. Also, even though this book deals with personnel selection, the proposed model is not an advocate for better selection as the preferred strategy for performance improvement.

4. The model does not "ignore" the fact that individual performance can be influenced by better tools, equipment, software, and so on. For example, the performance of air traffic controllers on a job-specific task factor called "using rules of separation efficiently" may be greatly influenced by technological changes in computerized radar displays. In fact, the change in technology may change the declarative knowledge, procedural knowledge and skill, and/or motivation requirements, which may in turn require different abilities or training courses. However, for purposes of measurement, the model simply stipulates that performance differences produced by differences in technology (for example, some job holders having better software than others) should not be confounded with performance differences produced by individual differences in declarative knowledge, procedural knowledge and skill, and motivation. That is, such external causes should be held constant; otherwise, the results of selection procedure validation, training program evaluation, or the results of evaluating any other such human resource intervention will be contaminated and misleading.

5. A similar argument can be made about effects of "constraints" on performance (Bernardin, 1989; O'Conner, Peters,

Weekley, Frank, & Erenkrantz, 1984). Constraints exert their influence on performance through their influence on motivation. If the constraints are changed, performance and the antecedents of the performance determinants may change. However, for purposes of selection research or training evaluation, the constraints must not be allowed to operate differentially within the research sample. Also, if the constraints on the research sample are not the same as those that will operate in the future population of incumbents, then corrections for the restriction must be used when population validity is estimated (Campbell, 1990a).

6. If performance really is multidimensional, then it is also true that we must face the issue of how to combine factor scores for purposes of making specific decisions. For example, academic departments seem to select faculty almost exclusively on the basis of factor 1, with some consideration for factor 3. Perhaps it is no wonder that they are often such contentious and strangely managed places. In general, individuals have different patterns of strengths and weaknesses. How are they to be compared? The specification of the combinational rules is again an expert judgment regarding how the goals of the organization can best be served. The appropriate procedure for estimating the parameter values is a research question.

7. This chapter is not about doing performance appraisal. There is a fundamental difference between measuring performance for purposes of evaluating a personnel management strategy (for example, selection or training) and making an operational decision about an individual (terminate, promote, grant a salary increase, and so on). For quite legitimate reasons, the same method can produce a very different distribution of scores for the same sample of people as a function of the two different objectives. Performance appraisal is certainly no less important than training evaluation or criterion-related test validation, but it is not the topic of this chapter. However, the model described here does imply that using operational performance appraisals as criteria in selection research is a dangerous thing to do. For similar reasons, such methods as management by objectives (MBO) and the conceptual frameworks of McGregor (1960), Odiorne (1965), and Deming (1986) are not discussed here.

Rather than ruminating over false arguments such as the above, readers would benefit more by taking issue with the content taxonomy, the characterization of the measurement issues, or the basic definition of performance itself. Our basic hope is that I/O psychology can come to some agreement about the meaning of performance so that future research can be designed, implemented, and interpreted within a common framework, broad though it may be.

Conclusion

Let us close by arguing that the greatest benefit of a model like the one proposed here is its value for guiding the choice of the dependent variable in a wide variety of research efforts, for structuring the reporting of research, and for organizing the accumulation of research results. Although black box correlations may be useful in some operational sense, it does very little for our collective understanding to say that general mental ability predicts promotion rate in most organizations, or that training in self-management lowers labor costs, or that the average correlation of assessment center evaluations with supervisor ratings of overall performance is .40. What will have far more value is to attach research results to a known latent structure for the dependent variable. Then we will no longer ask general questions such as How valid is this test? or How good is this or that type of training program? which is the way research results have historically been summarized and which focuses on the independent variable with no reference to its purpose. If the dependent variable is the variable of real interest, then the research questions should be of the form, What are the best strategies for improving peer leadership and support in organizations? or What must new entrants to the high-technology workplace know and be able to do to be successful on factors 1 and 5 in a particular population of jobs?

Finally, the mismatch of supply and demand for declarative knowledge, procedural knowledge and skill, and motivation in the labor force that will loom ever larger as we move into the twenty-first century (Johnston & Packer, 1987) is a problem of

enormous if not catastrophic magnitude, and is very much in the domain of applied psychology. Without a better conceptual structure for performance itself, we will attenuate our contributions to solving this problem. It would be better if that did not happen.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102,* 3–27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288–318.

Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). New York: W. H. Freeman.

Bass, B. M. (1990). *Bass and Stogdill's handbook of leadership: Theory, research, and managerial applications* (3rd ed.). New York: Free Press.

Bernardin, H. J. (1989). Increasing the accuracy of performance measurement: A proposed solution to erroneous attributions. *Human Resource Planning, 12*(3), 239–250.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology, 40,* 135–144.

Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L. A. (1991, April). *The role of early supervisory experience in supervisory performance.* Paper presented at the meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Campbell, C. H., Ford, P., Rumsey, M. G., Pulakos, E. D., Borman, W. C., Felker, D. B., de Vera, M. V., & Riegelhaupt, B. J. (1990). Development of multiple job performance measures in a representative sample of jobs. *Personnel Psychology, 43*(2), 277–300.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Campbell, J. P. (1990a). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, 2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P. (1990b). The role of theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, 2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P. (Ed.). (1991). *Building the career force—First annual report.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. (1973). The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology, 57,* 15–22.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43*(2), 313–333.

Campbell, J. P., & Zook, L. M. (Eds.). (1990). *Improving the selection, classification, and utilization of army enlisted personnel: Final report on Project A.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Cooper, W. H. (1981a). Ubiquitous halo. *Psychological Bulletin, 49,* 499–520.

Cooper, W. H. (1981b). Conceptual similarity as a source of illusory halo in job performance ratings. *Journal of Applied Psychology, 66,* 302–307.

Deming, W. E. (1986). *Out of the crisis.* Cambridge, MA: Massachusetts Institute of Technology Center for Advanced Engineering Studies (MIT-CAES).

DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance, 33,* 360–396.

Dunnette, M. D. (1963). A note on *the* criterion. *Journal of Applied Psychology, 47,* 251–254.

Fleishman, E. A. (1972). On the relationship between abilities, learning, and human performance. *American Psychologist, 27,* 1017–1032.

Gagne, R. M., Briggs, L. J., & Wager, W. W. (1988). *Principles of instructional design* (3rd ed.). New York: Holt, Rinehart & Winston.

Green, B. F., & Wigdor, A. K. (Eds.). (1988). *Measuring job competency: Report of the Committee on the Performance of Military Personnel. National Research Council.* Washington, DC: National Academy Press.

Hedberg, P. R. (1989). *Individual perceptions of performance feedback in a work setting.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. Ackerman, R. Sternberg, & R. Glaser (Eds.), *Learning and individual differences.* New York: Freeman.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. P., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581–595.

Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisory ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory.* Hillsdale, NJ: Erlbaum.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340–362.

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. In B. M. Staw & L. L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 5, pp. 141–197). Greenwich, CT: JAI Press.

James, L. R. (1973). Criterion models and construct validity for criteria. *Psychological Bulletin, 80,* 75–83.

Jenkins, J. G. (1946). Validity for what? *Journal of Consulting Psychology, 10,* 93–98.

Johnston, W. B., & Packer, A. E. (1987). *Workforce 2000 (The "executive" summary).* Indianapolis, IN: Hudson Institute.

Kanfer, R. (1990). Motivation theory and industrial/organiza-

tional psychology. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (2nd ed. Vol. 1). Palo Alto, CA: Consulting Psychologists Press.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative-aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74,* 657–690.

Kyllonen, P. C. (1985). *Dimensions of information processing speed* (Report No. AFT-TP-84-56). Brooks AFB, TX: Air Force Human Resource Laboratory, Manpower & Personnel Division.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87,* 72–107.

Lohman, D. F. (1989). Estimating individual differences in information processing using speed-accuracy models. In R. Kanfer, P. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation, and methodology: The Minnesota symposium on learning and individual differences.* Hillsdale, NJ: Erlbaum.

Mahoney, T. A. (1988). Productivity defined: The relativity of efficiency, effectiveness, and change. In J. P. Campbell & R. J. Campbell (Eds.), *Productivity in organizations.* San Francisco: Jossey-Bass.

McCloy, R. A. (1990). *A new model of job performance: An integration of measurement, prediction, and theory.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

McGregor, D. (1960). *The human side of enterprise.* New York: McGraw-Hill.

McLagan, P. A. (1988). Flexible job models: A productivity strategy for the information age. In J. P. Campbell & R. J. Campbell (Eds.), *Productivity in organizations.* San Francisco: Jossey-Bass.

Murphy, K., & Balzer, W. (1989). Rating errors and rating accuracy. *Journal of Applied Psychology, 74,* 619–624.

Nagle, B. F. (1953). Criterion development. *Personnel Psychology, 6,* 271–289.

Nathan, B. R., & Alexander, R. A. (1988). A comparison of criteria for test validation: A meta analytic investigation. *Personnel Psychology, 41,* 517–536.

Nathan, B. R., & Tippins, N. (1990). The consequences of halo

"error" in performance ratings: A field study of the moderating effect of halo on test validation results. *Journal of Applied Psychology, 75,* 290–296.

Nissen, M. J., & Bullmer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1–32.

O'Conner, E. J., Peters, L. H., Weekley, J., Frank, B., & Erenkrantz, B. (1984). Situational constraints and effects on performance, affective reactions, and turnover: A field replication and extension. *Journal of Applied Psychology, 69,* 663–672.

Odiorne, G. S. (1965). *Management decision by objectives.* Englewood Cliffs, NJ: Prentice-Hall.

Oppler, S. H. (1990). *Three methodological approaches to the investigation of subgroup bias in performance measurement.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Ostroff, C., & Ford, J. K. (1989). Assessing training needs: Critical levels of analysis. In I. L. Goldstein (Ed.), *Training and development in organizations.* San Francisco: Jossey-Bass.

Patterson, C. H. (1946). On the problem of the criterion in prediction studies. *Journal of Consulting Psychology, 10,* 277–280.

Pritchard, R. D., Jones, S. D., Roth, P. L., Steubing, K. K., & Ekeberg, S. E. (1988). Effects of group feedback, goal setting, and incentives on organizational productivity. [Monograph]. *Journal of Applied Psychology, 73,* 337–358.

Pritchard, R. D., & Roth, P. G. (1991). Accounting for nonlinear utility functions in composite measures of utility and performance. *Organizational Behavior and Human Decision Processes, 50,* 341–359.

Rousseau, D. M. (1989). *Psychological contracts in recruitment.* Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Boston.

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88,* 413–428.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73,* 482–486.

Sadacca, R., Campbell, J. P., DiFazio, A. S., Schultz, S. R.,

& White, L. A. (1990). Scaling performance utility to enhance selection/classification decisions. *Personnel Psychology, 43,* 367-378.

Sager, C. E. (1990). *A component model of halo: Peer and supervisory ratings of job performance.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Schein, E. H. (1970). *Organizational psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37,* 407-422.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: 2. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques.* New York: Wiley.

Toops, H. A. (1944). The criterion. *Educational and Psychological Measurement, 4,* 271-297.

Vance, R. J., MacCallum, R. C., Coovert, M. D., & Hedge, J. W. (1988). Construct validity of multiple job performance measures using confirmatory factor analysis. *Journal of Applied Psychology, 73,* 74-80.

Wallace, S. R. (1965). Criteria for what? *American Psychologist, 20,* 411-417.

Yukl, G. (1987, August). *A new taxonomy for integrating diverse perspectives on managerial behavior.* Paper presented at the meeting of the American Academy of Management, New Orleans, LA.

3

Expanding the Criterion Domain to Include Elements of Contextual Performance

Walter C. Borman, Stephan J. Motowidlo

Individuals contribute to organizational effectiveness in ways that go beyond the activities that comprise their "jobs." They can either help or hinder efforts to accomplish organizational goals by doing many things that are not directly related to their main task functions but are important because they shape the organizational, social, and psychological context that serves as the critical catalyst for task activities and processes. These contextual activities are sometimes represented in criterion dimensions of individual job performance, and they are sometimes implicitly recognized as important for organizational effectiveness. They are frequently ignored, however, and as a result, are not included, explicitly or implicitly, in assessments of individual job performance to be used as criteria.

In this chapter we argue that selection criteria should embrace a domain of organizational behavior broader than just task activities; they should also include contextual activities. First, we discuss contextual performance in general and show how it differs from task performance. Then we discuss four streams of research that illustrate aspects of contextual perfor-

Note: The Authors thank Irv Goldstein, Neal Schmitt, Ben Schneider, and Dennis Organ for commenting on an earlier draft of this chapter.