A Review and Extension of Current Models of Dynamic Criteria

Debra Steele-Johnson*, Hobart G. Osburn and Kalen F. Pieper

An important issue in personnel selection and test validation has been the nature of performance criteria and more specifically the existence of dynamic criteria. There is a continuing debate regarding the extent to which performance and validity coefficients remain stable over time. We examine research within work, laboratory, and academic settings and evaluate existing models of dynamic criteria. Building on previous models, we propose an integrative model of dynamic criteria that identifies important issues for ability and performance constructs and discusses how variables related to the task, job, and organization can affect the temporal stability of criterion performance and the ability-performance relationship.

Introduction

The study of personnel selection and test validation has a long and rich history (see Borman 1991; Guion 1991; Hilgard 1987). A important issue in personnel selection and test validation is the stability over time of performance criteria and more specifically the temporal stability of relationships between predictors and performance criteria. This issue has been labelled dynamic criteria. The study and understanding of dynamic criteria are important because an assumption of stable relationships between predictors and criteria is inherent in every personnel decision in organizations (Hanges, Schneider and Niles 1990). In general, the traditional selection paradigm assumes that the rank order of individuals on performance measures remains stable over time and, further, that validity coefficients obtained from the original validation study are also stable over time. Therefore, questioning the assumption that validity coefficients are stable is of paramount importance practically because changing validity coefficients affect the expected utility of a selection programme and theoretically because questioning the assumption can lead us to learn more about the nature of human performance (Schmidt, Ones and Hunter 1992).

Consequently, our objectives in this article are first, to review the literature on sources of variation in predictive validity estimates when performance is measured at different points in time, and, second, to propose an extended model that focuses on sources of variation in ability-performance relationships that can be used to guide future research. Specifically, we propose a model describing the role of task, job, and organization level variables in criterion performance and ability-performance relationships and the role of skill acquisition or work experience in those effects.

In the following discussion we examine the assumption of stability of performance and validity coefficients and address the implications of instability. First, we examine key examples of evidence that support or contradict the existence of dynamic criteria. An important feature of this review is that we examine research regarding dynamic criteria within work, laboratory, and academic settings to determine whether similar results are obtained in these settings. Second, we examine existing models of dynamic criteria, the conditions they identify that are likely to affect the stability of criteria, and how adequate they are in explaining the phenomenon. Finally, building on previous models, we propose an integrative model of dynamic criteria. Our model identifies important issues for ability and performance constructs and discusses how task, job, organizational, and learning-related variables are likely to affect criterion performance and ability-performance relationships. We extend previous research (e.g., Deadrick and Madison 1990) and models (e.g., Hesketh and Robertson 1993) by using a more fine-grained approach to evaluating the research on dynamic criteria (i.e., within settings) and integrating and extending previous models to offer an integrative framework for understanding dynamic criteria.

Origin and Definition of the Issue

The assumption of stable performance and stable validity coefficients has been questioned for
some time. Researchers have noted that the performance of incumbents changes as they learn and develop on the job (Ghiselli 1956) and that job performance requirements may change over time (Jenkins 1946). As a result, researchers have questioned whether tests predict performance over the entire employment period (Worbois 1951; see also Bass 1962; Ghiselli and Haire 1960). Although substantial research has examined the stability of performance and validity coefficients, methodological problems (e.g., Barrett, Caldwell and Alexander 1985; Bechtoldt 1962; Humphreys 1960) and conflicting research results (e.g., Barrett, Alexander and Doverspike 1992; Henry, Hulin and Noon 1990) have prevented a resolution of the issue.

Moreover, models of performance change (e.g., changing task model, changing subject model) have been developed, but these models have been unable to adequately explain dynamic criteria (see Alvares and Hulin 1972, for a discussion). Researchers have made very limited predictions about when dynamic criteria are likely to occur and how validity coefficients are likely to change (for an exception, see Hesketh and Robertson’s process model [1993]). As a result, research on dynamic criteria appears fragmented and sometimes contradictory. We propose a model that attempts to explain dynamic criteria more adequately and makes specific predictions about the nature and occurrence of dynamic criteria.

The term dynamic criteria has been used to describe two separate but interrelated processes (Barrett, Caldwell and Alexander 1985). One use of the term refers to a change in the rank order of subjects’ performance over time and involves examining the stability of a single performance measure across two or more measurement occasions. Typically, systematic changes in performance over time produce a pattern of correlations in which the correlations between performance measures systematically decrease with increases in the time interval between measurements. This is referred to as the simplex pattern which will be discussed more fully in a later section. One should note that there is little disagreement that there are substantial changes over time in the rank order of individuals on most tasks.

The term also refers to changes over time in validity coefficients between predictors and criteria. Typically, the predictor is administered only once so that time-dependent change in validity coefficients must be due to changes in subsequent performance. There are three potential reasons for systematic changes in the performance of subjects over time: (1) subjects might change the way they perform the task (Woodrow 1938a, 1938b); (2) the knowledge and ability requirements needed to perform the task might change (Fleishman and Hempel 1954, 1955); or (3) the knowledge and abilities of subjects might change (Alvares and Hulin 1972, 1973). We suggest that the term, dynamic criteria, has tended to narrow the focus of research by focusing on the criteria rather than the causes of changes in performance, and for this reason has retarded developments in this field. The crucial process in dynamic criteria is change in predictor-criterion relationships over time, and as shown in this article, these relationships may be influenced by a wide variety of important variables that affect performance.

**Key Examples of Research in Work, Laboratory and Academic Settings**

Key examples of research from work, laboratory, and academic settings suggest that the existence of dynamic criteria has not been disproved and highlight many of the issues relating to dynamic criteria, some of which remain unresolved.

**Research in Work Settings**

Two frequently cited early studies claiming evidence of dynamic criteria were Ghiselli and Haire (1960) and Bass (1962). Unfortunately, as others have noted (e.g., Deadrick and Madigan 1990), these studies were flawed, making it difficult to draw conclusions from them. For example, much of Ghiselli and Haire’s (1960) results were probably mostly due to sampling error. Additionally, Bass’ (1962) validities were low, inconsistent, and biased by restriction of range. An early study by Worbois (1951) claimed that validities were stable. Unfortunately, there were multiple problems with this study including differences in criteria and capitalizing on sampling error.

In contrast to these earlier studies, a well-designed study by Deadrick and Madigan (1990) found a slight increase in the validity of a cognitive test in predicting output of sewing machine operators over a six-month period. Further, other research (Rothe 1946, 1947, 1951, 1978; Rothe and Nye 1958, 1959) has focused on the consistency of individuals’ performance on successive days or in successive weeks (rather than changes in predictor-criterion relationships). Rothe and his associates concluded that performance is moderately consistent, but organizational variables such as financial incentives can increase performance consistency. On the other hand, Rambo, Chomiak, and Price (1983) have suggested that other factors such as task complexity and contextual factors might decrease the stability of performance.
Research in Laboratory Settings

Woodrow (1938a, 1938b) and Fleishman and his associates (e.g., Fleishman 1953; Fleishman and Hempel 1954, 1955) have been important contributors to laboratory research addressing dynamic criteria. Using factor-analytic methodology, Fleishman and his associates (e.g., Fleishman 1953; Fleishman and Hempel 1954, 1955) found that the importance of different ability factors varied with practice on factorially complex psychomotor tests, i.e., rudder control tasks. For example, Fleishman and Hempel (1954) found that Spatial Relations, Visualization, and Mechanical Experience were primarily involved in task performance early in task practice, but Psychomotor Coordination, Rate of Movement, and a factor that was specific to the task itself were most important late in practice. The central finding that the pattern of abilities determining proficiency in psychomotor performance tasks may change as a function of practice was confirmed using a variety of laboratory tasks, including a Discrimination Reaction Time Test (Fleishman and Hempel 1955), a Rotary Pursuit Test (Fleishman 1960), a Two-Hand Coordination task (Fleishman and Rich 1963), and a Morse code task which represented an extension of the previous lab results to an actual job setting (Fleishman and Fruchter 1960).

Fleishman and Hempel (1955) suggested that once the practice task is learned to a sufficiently high degree, some abilities become less relevant and other combinations of abilities become more important contributors to performance. This explanation is consistent with Woodrow’s (1938a, 1938b) focus on the changing nature of the task during practice. As will be discussed in a later section, these ideas are also consistent with the concepts of automatic and controlled processing (Schneider and Shiffrin 1977; Shiffrin and Schneider 1977).

Although Fleishman made great strides in the study of individual differences, his work is not without its critics (e.g., Barrett, Caldwell and Alexander 1989; Bechtoldt 1962). Humphreys (1960) contended that because adjacent trial correlations are more highly correlated than trials further apart (the simplex pattern), exploratory factor analysis was inappropriate. Similarly, Bechtoldt (1962) noted that using factor analysis to explore the effects of independent variables on dependent variables is inappropriate both experimentally and statistically.

Various re-analyses have been conducted to address these concerns. For example, Bechtoldt’s (1962) re-analysis of Fleishman and Hempel’s (1954, 1955) data using multiple regression indicated that reaction time and motor factors predicted performance over time, but the factors were not the same as those described by Fleishman and Hempel. Additionally, Barrett et al.’s (1989) re-analysis using the Dunn and Clark (1969, 1971) test for correlations revealed very few significant differences between adjacent pairs of correlations. However, this test is very conservative with low statistical power compared to other possible tests (Austin et al. 1989).

More importantly, when studying longitudinal changes in correlations, it is the systematic pattern of correlations that is important — not differences between pairs of correlations.

More recent work (Fleishman and Mumford 1989) has addressed many of the concerns expressed by other researchers. For example, in some of Fleishman’s later studies (see Fleishman and Mumford 1989, for example), the reference test battery was factored separately, and then the loadings of the trial stages were projected on these factors using Dwyer’s extension technique (Dwyer 1937). Finally, although re-analyses by Barrett, Morris and Alexander (1993) of Fleishman’s and Hempel’s work (i.e., Fleishman 1960; Fleishman and Hempel 1954, 1955) produced a different pattern of results than those described by Fleishman, the re-analyses did not eliminate the central effects of changing ability correlations with practice on a criterion task.

Research in Academic Settings

Humphreys and his associates (e.g., Humphreys 1968; Humphreys and Taber 1973; Powers 1982; Winterbottom, Pitcher and Miller 1976) have examined dynamic criteria in the context of academic tasks. For example, Humphreys (1968) examined the stability of college grade point averages (GPAs) for approximately 1600 students across eight semesters of college work. His results provided strong evidence of a systematic change in performance such that performance measures taken close together were much higher than performance measures more widely separated (the simplex pattern). In addition, Humphreys (1968) showed that the validity of the ACT test and high school rank as predictors of semester GPA declined steadily over the eight semesters. Further, when the correlations were corrected for unreliability in the GPAs, the true score correlations were much higher but showed the same pattern. Similarly, Humphreys observed the same pattern in subsequent research (Humphreys and Taber 1973; Lin and Humphreys 1977).

Humphreys’ results are compelling because the results were based on a well-designed, large sample study in which the correlations were corrected for restriction of range. Moreover, other research examining predictive validities has produced similar results, showing that
Humphreys' finding is robust. For example, Juola (1966) observed an almost identical drop-off in validity for the College Qualification Test (CQT) in predicting undergraduate GPA. Also, Lin and Humphreys (1977) found a similar decline in validity for the Graduate Record Examination (GRE) in predicting graduate GPA, and studies of the Law School Admission Test (LSAT) have shown a decline in validity for the LSAT in predicting law school GPA (Powers 1982; Winterbottom et al. 1976).

However, these results have been found primarily when GPAs were computed for individual semesters. Studies using cumulative GPAs have failed to find declining validities in the later semesters of undergraduate work. For example, Hills, Bush, and Klock (1964) found comparable multiple correlations using the Scholastic Aptitude Test (SAT) and high school GPA for predicting cumulative GPAs at the end of the sophomore year and the senior year. Also, Mauger and Kolmodin (1975) found no decline in validity of the SAT to predict cumulative GPA at the time the student left college either due to graduation or other reasons, compared with the prediction of freshman GPA.

Other research has examined whether postdictive validities show similar declines. For example, researchers have observed declining validities using GRE scores for postdicting undergraduate GPA from the freshman through senior years (Humphreys and Taber 1973; Lin and Humphreys 1977). However, other research has failed to reveal declines in validities using the GRE Advanced Test for postdicting undergraduate GPA (Humphreys and Taber 1973) and for the LSAT for postdicting law school GPA (Winterbottom et al. 1976). One possible explanation is that the LSAT and GRE Advanced Test might sample achievement to a greater extent than the SAT, ACT or GRE verbal and quantitative tests. Thus, one would expect LSATs and GRE Advanced Tests to predict more consistently across years in school when the LSAT or GRE Advanced Test is administered after the maximum opportunity is provided to obtain the sampled skills, i.e., at the completion of law school or undergraduate education.

Humphreys (1968) argued strongly for the possibility that changes in students' abilities were at least partly responsible for the decline in validity coefficients, i.e., that subjects were developing intellectually at different rates which would explain the decline in validity of the predictors. Humphreys' argument is consistent with Alvares and Hulin's (1972) description of a changing subject model (see below). However, Humphreys' (1968) hypothesis that changing abilities account for the change in validities was not supported by Humphreys and Taber's (1973) and Lin and Humphreys' (1977) results. Indeed, GRE scores postdicted freshman GPAs better than senior GPAs. An earlier study by Lunneborg and Lunneborg (1970) also found the same pattern for both predictive and postdictive validities. These studies essentially ruled out the hypothesis that substantial changes in the construct measured by the predictor were responsible for these findings. Clearly, in light of these data, a changing abilities explanation is implausible or at least cannot be the whole story.

We offer three alternative hypotheses: change in the criterion, first year intellectual challenge, and flexibility in course choice. With respect to change in the criterion, large sample studies have shown that the validity of the LSAT declines slightly in predicting second and third year law school grades (Carlson and Werts 1976; Lin and Humphreys 1977; Powers 1982). One possible explanation is that the curriculum in the second and third years might be changing in such a way as to emphasize skills other than those measured by the LSAT. An alternative hypothesis is that the first year in a new academic situation might present more of an intellectual challenge compared with later years (Lin and Humphreys 1977). Therefore, the LSAT which measures 'ability to confront new situations' (Powers 1982) might be measuring abilities that are more salient to the first year challenge. The strongest support for the intellectual challenge hypothesis is the fact that postdictive studies show (with minor exceptions) essentially the same pattern as do predictive studies, i.e., the validity of college aptitude tests is highest for first year grades and steadily declines in subsequent semesters. Also, the fact that the validity of undergraduate GPAs does not decline over the three years of law school provides indirect support for this hypothesis (Carlson and Werts 1976; Winterbottom et al. 1976; Johnson and Olsen 1976; Powers 1982). A third hypothesis suggested by Powers (1982) is that the decline in validity is due to changes in the rank order of students brought about by their academic decisions. Undergraduate students enjoy considerable flexibility in choosing courses and instructors, in changing majors, and in tailoring the academic situation to their particular abilities and interests. The fact that the validity of the ACT and the CQT in predicting undergraduate GPA declines more precipitously than the validity of the LSAT in predicting GPA in a more structured law school environment is consistent with this hypothesis. In addition, a study of the validity of the SAT in predicting GPA at the US Military Academy shows no decline in validity over the four year period (Butler and McCauley 1987). The most reasonable explanation of this result is the very structured academic environment at the Military Academy that allows a student almost no flexibility in tailoring the academic situation to his/her abilities and interests.
Each of these alternative hypotheses is consistent with what Alves and Hulin (1972) have referred to as the changing-task model or the Fleishman-Woodrow model, i.e., the relative importance of different abilities required to perform the task might change over time either because of extended practice on a task or because of actual changes in the task to be performed (see below).

Summary

In summary, research relevant to dynamic criteria has been conducted in work, laboratory, and academic settings. Drawing conclusions from the research in work settings is difficult due to methodological problems in many of the studies. However, stronger conclusions can be drawn from studies using laboratory tasks. This research has shown that (a) correlations among performance measures after successive practice trials inevitably show a simplex pattern of higher correlations among adjacent pairs and lower correlations between measures taken farther apart; and (b) whereas the predictive validity of most abilities declines with practice, the validity of some psychomotor abilities increases with practice on the criterion task. Similarly, strong conclusions can be drawn from research in academic settings. As was true in laboratory studies, the correlations among GPAs measured over successive time periods show a definite simplex pattern (i.e., performance measures taken closer in time show higher correlations than measures that are more widely separated). Second, the predictive validity of cognitive ability and high school grade point average decline steadily from the first to the last semesters of college work. Also, postdictive studies of GPAs from the GRE and similar tests show the same pattern of declining validities as has been observed in predictive studies – data that are consistent with the changing-task model but hard to reconcile with a hypothesis of changing abilities as a result of intellectual development.

The Simplex Pattern

The simplex pattern has often been cited as evidence of dynamic criteria, highlighting the need for a clearer understanding of the simplex, possible explanations for its existence, and the extent to which it provides evidence of dynamic criteria. Although a few investigators (e.g., Barrett and Alexander 1989) have disputed the existence of the simplex, a simplex pattern or a simplex-like pattern is ubiquitous among repeated measures of performance, including job performance (Deadrick and Madigan 1990; Hofmann, Jacobs and Baratta 1993), academic performance (Humphreys 1968), teaching performance (Hanges et al. 1990) and sport performance (Hofmann, Jacobs and Gerras 1992).

Kincaid (1925) and Perl (1934) were among the first to comment on the occurrence of a simplex pattern, noting a tendency of experiments with practice periods to show lower correlations between initial and final ability measures, compared with measures closely spaced in time. Jones (1962), in an explanation of the simplex pattern in the context of skill acquisition, posited that a simplification process underlies skill acquisition. However, Guttman (1955) most clearly addressed this issue and proposed the additive linear model as a theoretical explanation for the observed simplex pattern of correlations. Specifically he proposed a quantitative simplex model in which the difference between scores at Time j and Time k is uncorrelated with the difference between scores at Time k and Time l. One way to obtain this condition is to obtain the next score in the sequence by adding a score uncorrelated with the previous score. The additive linear model says that given a sequence of scores $s_1, s_2, s_3, s_k, s_{l+1}$, the difference between any $s_k$ and any of its predecessors in the sequence is uncorrelated with the difference between this same $s_k$ and any of its successors in the sequence. This is equivalent to adding to $s_1$ an increment uncorrelated with $s_1$ to get $s_2$, and adding to $s_2$ an increment uncorrelated with $s_2$ to get $s_3$, and so on. This condition will generate a 'perfect simplex'. It should be noted, however, that whereas the Guttman simplex model fits the simplex pattern of correlations quite well, other growth models can also fit the simplex pattern equally well (Rogosa and Willett 1985).

The concept of temporal bias offers a substantive explanation of the simplex that is entirely compatible with Guttman's additive model. Temporal bias refers to more or less random environmental and internal events that affect individuals and their rank order on the attribute being measured. These temporary events need not be the same for different individuals. All that is needed is that such events have a temporary effect on the scores of individuals on the attribute being measured. Different events may impact different individuals. We posit that the effects of such environmental or internal events persist across periods closer in time but gradually decay across time periods that are widely separated. Thus, the effects of these events contribute more strongly to the correlation between measurements at time periods that are relatively close together and make a diminishing contribution to the correlation between measurements taken at time periods that are widely separated. We assume...
that there is temporal bias at every time period measurement, and that temporal bias is, first, independent of the attribute (e.g., a specific work skill) being measured; and, second, independent of the temporal bias that occurs at more remote time periods. Temporal bias occurs most strongly in situations in which individuals have some measure of control over their environment, i.e., when there is reciprocal causation in which the environment affects the individual and is in turn affected by individual decision-making.

The presence of a simplex or simplex-like pattern does not necessarily indicate progressive decrements in the validity of a predictor (Schmidt 1993), and thus we must be cautious in drawing inferences that predictor-criterion relationships are changing based on evidence of a simplex-like pattern for correlations among repeated performance measures. The simplex pattern is seen as an artefact due to the effects of temporal bias that are essentially unpredictable over the long term. However, a lasting general component could be more, less or equally predictable due to learning, context factors, etc.

Our formulation accounts for several examples of the simplex. For instance, the simplex pattern is ubiquitous for very different types of variables, suggesting that the simplex pattern may be due to artefacts resulting from repeated measurements rather than the nature of the variables studied. Joreskog and Sorbom (1989) showed that GPA data from Humphreys (1968) could be modelled by a quasi-simplex in which cognitive ability causes performance in the first semester and has only an indirect effect on performance in subsequent semesters. Postdictive and predictive studies of scholastic ability show the same pattern, i.e., in both instances initial performance is predicted better than later performance. Thus, although there are changes in performance over time, these changes are not due to changes in scholastic ability but must be due to other factors. Finally, when individuals are very similar and the environment is very homogeneous, the simplex pattern is reduced or non-existent (Butler and McCauley 1987).

Previous Models of Dynamic Criteria

The observation that abilities might be unstable was introduced decades ago. Humphreys (1968) suggested that declines in validity coefficients over time can be accounted for at least in part by changes in individuals’ abilities. His suggestion and critics’ alternate explanations are captured in the two original models proposed to explain dynamic criteria: the changing-subject and changing-task models. Other elaborations on these models have followed.

The changing-subject model emerged from various researchers, most notably Adams (1957) who suggested that an individual’s abilities can change with task practice. This hypothesis was in a framework of a non-traditional conceptualization of abilities, i.e., that abilities reflect an individual’s current repertoire of responses. Others (e.g., Humphreys 1960) have presented views consistent with Adams, arguing for the usefulness of conceptualizing abilities in terms of one’s current repertoire of available responses.

The changing-subject model can be expressed mathematically (Alvares and Hulin 1972):

$$x_{ij} = \sum_k (a_k y_{ik}) + e_{ij} + s_j$$

where $x_{ij}$ is the deviation score of the $i$th person on the $j$th trial, $a_k$ is the importance of the $k$th common ability factor on the $j$th trial, $y_{ik}$ is the amount of the $k$th common ability possessed by the $i$th individual, $e_{ij}$ is the error of measurement for the $i$th individual on the $j$th trial, and $s_j$ is a factor specific to the $j$th trial. An important limitation of the model is that the covariance between ability component scores is assumed to be zero. The changing-subject model asserts that $y_{ik}$, the amount of a given ability possessed by a given individual, changes with task practice, learning or experience on the criterion task. This is in contrast to the changing-task model (see below) that asserts that $a_k$, the weight associated with a given ability, changes with practice, learning or experience.

The changing-subject model offers a theoretical explanation for temporal changes in validities and performance. However, there is not much solid evidence evaluating the model because studies usually fail to obtain post-performance ability measures (e.g., Humphreys 1960). Moreover, studies of academic performance that have measured post-performance ability suggest that the changing abilities explanation is implausible or at least not the whole story (Humphreys and Taber 1973; Lin and Humphreys 1977; Lunneborg and Lunneborg 1970). In addition, Dunham (1974) in a laboratory study and Alvares and Hulin (1973) in a flight training study found only minimal evidence for changes in ability as the result of practice on a criterion task.

The changing-task model, attributed to Woodrow (1938a) and Fleishman (1960, 1966), suggests that the importance of some abilities increases while the importance of other abilities decreases with continuing practice on a task. Studies show that general cognitive factors play a more important role in early practice trials and motor abilities become important in later practice trials (Fleishman 1960, 1966). Fleishman’s hypothesis is consistent with Jones’ (1966) suggestion...
that general abilities are more important early and that narrower and fewer abilities become important later in practice on a task. Additionally, Ackerman (e.g., Ackerman 1988, 1992, Ackerman and Kanfer 1993; Kanfer and Ackerman 1989, see below) has provided substantial evidence that the contribution of abilities to the prediction of performance changes over time.

More recently, other researchers have offered elaborations of the original two models. For example, Dunham (1974) has suggested that dynamic criteria reflect both changing tasks and changing subjects. Ackerman (1986, 1988) has extended the changing-task model by addressing the role of individual difference factors in skill acquisition and the role of task complexity and task consistency as moderators of ability-performance relationships. Similarly, Murphy’s (1989b) dynamic model of job performance is derived from a changing-task model, focusing on how changes in the nature of the job influence the relative importance of different individual difference factors, suggesting that ability is more important in the transition stage, that is, when the job is changing, and that other factors (e.g., motivation, personality) increase in importance during the maintenance stage, that is, when the job is stable. Hanges et al. (1990) extended previous research by integrating Murphy’s (1989b) model with an interactionist perspective to explore the stability of performance during the maintenance stage, positing that behaviour is jointly determined by the person and the situation (Bowers 1973). Finally, Hesketh and Robertson (1993) have presented a more comprehensive model for selection research, suggesting that we can improve our selection processes by addressing a wider array of predictors (e.g., motivation and personality) and criteria (e.g., tenure, job satisfaction) and by addressing the role of task and context factors in these additional predictor-criterion relationships. Our model complements Hesketh and Robertson’s in that their model suggests that selection processes can be improved by expanding the range of factors we assess as predictors and criteria whereas our model suggests that we also can increase our understanding of selection processes by taking a more fine-grained approach, focusing on time-related changes in the ability-performance relationship. Thus, the model we propose below integrates factors and processes from previous models, offers a more comprehensive examination of the ability-performance relationship and factors affecting it, and provides an organizing framework for future research that considers the role of different factors in ability-performance and other predictor-criterion relationships.

An Integrative Model of Dynamic Criteria

Current models (Ackerman 1986, 1988; Hanges et al. 1990; Hesketh and Robertson 1993; Murphy 1989b) have addressed a number of factors likely to play a role in dynamic criteria, including task characteristics, personality, motivation, and contextual variables. Our model integrates previous models and provides a framework to guide future research by focusing on (a) the nature of abilities and performance; and (b) the role of task, job, and organizational variables in ability-performance relationships and how those effects are influenced by learning-related variables (see Figure 1). We begin by discussing the nature of cognitive abilities and performance and how the observed relationships between them is influenced by the measures we use. Then, we discuss the nature of tasks, jobs, and organizational variables, their role in the ability-performance relationship, and the influence of skill acquisition and work experience. Although a comprehensive examination of the direct effects of task, job, and organizational factors on performance is beyond the scope of this article, we identify and briefly discuss direct effects where appropriate. We focus in more detail on the research questions more directly related to dynamic criteria: how task, job and organizational factors moderate relationships between ability and performance.

Nature of Performance

As Jenkins (1946, p. 93) commented, ‘Psychologists in general tended to accept the tacit assumption that criteria were given of God or just to be found lying about.’ However, researchers (Campbell 1990; Campbell, Gasser and Oswald 1996; Campbell, McClay, Oppler, and Sager 1993; Murphy 1989a, 1990) have focused recently on identifying and explicitly assessing latent constructs underlying criterion performance. We examine performance constructs and distinctions between task and contextual factors, noting that they might be differentially predicted by abilities. Further, we discuss temporal instability in performance. Finally, we address how operationalizations of the performance construct, i.e., a performance measure, can affect interpretations of the ability-performance relationship. Although research examining performance on laboratory tasks or academic performance involve similar constructs, we discuss issues primarily in relation to work performance.
The Performance Construct

Research on the construct of performance has raised three important issues. First, as researchers have noted (Murphy 1989a, 1990; Campbell et al. 1996), there are few definitions of the performance construct in the literature. Murphy (1989a, 1990) defined the performance domain as 'the set of behaviours that are relevant to the goal of the organization or the organizational unit in which a person works'. Second, terms such as 'overall performance' imply that performance can be meaningfully viewed in economic terms as the sum total of an employee's contribution to the goals of the organization. Whereas the concept of overall job performance is useful in some contexts, most investigators have concluded that performance must be viewed as a multidimensional construct (e.g., Campbell et al. 1993). Finally, job performance often has been equated with task performance. In contrast, Murphy (1989a) has suggested that employees perform and are measured on many non-task behaviours (interpersonal activities, absenteeism, etc.). Thus, the construct of job performance should be defined more broadly than task proficiency.

Murphy (1989a, 1990) identified four broad clusters of behaviours: task-oriented behaviours, interpersonally-oriented behaviours, down-time behaviours (e.g., absenteeism), and destructive/hazardous behaviours. Subsequently, Campbell and co-workers (Campbell et al. 1993) proposed an eight factor model: 1. Job-specific task proficiency; 2. Non-job-specific task proficiency; 3. Written and oral communication; 4. Demonstrating effort; 5. Maintaining personal discipline; 6. Facilitating peer and team performance; 7. Supervision/Leadership; and 8. Management/Administration. Campbell et al. posited that these eight factors can describe most, if not all, jobs at a molar level although some dimensions might not be relevant for some types of jobs. It is important to note that only two or possibly three of Campbell's performance factors are directly related to cognitive abilities and skills.

Task Versus Contextual Performance

Borman and Motowidlo (1993) have provided another perspective by distinguishing between task and contextual performance. Task performance refers to the proficiency with which incumbents perform activities that contribute to an organization's technical core. Task performance activities are formally recognized as part of the job. In contrast, contextual performance includes prosocial behaviours that support the psychological and social environment of the organization but are often not considered to be formal requirements of the job. Investigators

---

Figure 1: An integrative model of dynamic criteria.
suggest that task proficiency will relate to individual differences in knowledge, skills, and abilities that are important to performing a given job, whereas contextual performance will likely be related to motivational and personality variables (Borman and Motowidlo 1993; Motowidlo and VanScotter 1994).

Implications of Performance Construct Operationalizations

Changes in individual or organizational factors can cause performance to be unstable, and it is important to have information about the stability of performance in a given context and to model factors affecting stability where possible. This is a key issue in dynamic criteria. However, it is important also to note that measures of performance play an important role in observed predictor-criterion relationships, specifically through their effects on the quality and characteristics of criteria. Further, Murphy and Shiarella (1997) have argued that performance is multidimensional and performance dimensions should be weighted according to relevant organizational goals such as individual versus team performance. To the extent that the weight of task proficiency diminishes relative to other performance dimensions such as prosocial behaviours, one can expect lower predictive validity for cognitive ability and higher predictive validity for other factors (e.g., personality, motivation). If the relative weight of task proficiency in job performance either increases or decreases (for example, because of practice, experience, changes in the job structure, or changes in organization policy) relative to other performance variables, changes in the predictive validity of cognitive ability would be expected.

Similarly, motivational factors can influence the measurement of performance, and hence, our interpretation of predictor-criterion relationships. For example, to the extent that the criterion measure is obtained unobtrusively or is part of the normal routine for the individual, one is more likely to assess typical performance. However, to the extent that the criterion measure is salient to the individual (e.g., work sample tests), one is more likely to obtain a measure of maximal (i.e., highest possible) performance. In selection, one probably wants to predict the expected typical level of performance of a prospective employee. Thus, researchers need to consider whether an employee is being validated against a measure of typical or maximal performance (see section on abilities below for other issues relating to typical versus maximal performance).

To summarize, in our model we focus primarily on how adequately criterion measures represent dimensions of the performance construct. New approaches (e.g., Campbell et al. 1993; Motowidlo and VanScotter 1994) to modelling the performance construct highlight the point that performance is multiply determined. Also, components of performance might differ in importance over time. Thus, what one chooses to assess in the criterion measure has important implications for observed ability-performance relationships.

Proposition 1: Measures of cognitive ability will be stronger predictors of criterion performance (a) when the performance measure is task proficiency; and (b) maximal performance, rather than typical performance is being assessed.

Nature of Abilities

Cognitive abilities have been a major focus in selection research, demonstrating substantial validities for diverse jobs (Hunter 1986; Hunter and Hunter 1984). Moreover, a comprehensive examination of the nature of all types of abilities (physical abilities, etc.) is clearly beyond the scope of an article of this length. Consequently, we focus our discussion on cognitive abilities, specifically, the constructs of general cognitive ability, aptitude, and achievement. We examine alternate conceptualizations of cognitive abilities, discuss distinctions between abilities and related constructs (e.g., achievement, skill), and examine evidence relating to their temporal stability.

Conceptualizations of Abilities

Cognitive ability is an acquired repertoire of mental capabilities involving the capacity to reason, manipulate symbolic materials, solve problems, think abstractly, visualize objects in space, perceive relationships, and comprehend complex ideas. Individual tasks reflecting cognitive ability differ in terms of the extent to which they represent the repertoire. Tasks that involve solving problems and thinking abstractly are more central to the concept whereas tasks involving short-term memory and relatively simple speeded checking tasks are more peripheral but still are part of the repertoire.

One way of organizing cognitive ability is in a hierarchical model such as suggested by Vernon (1950) and others (Ackerman 1989a; Burt 1949; Humphreys 1982; Lubinski and Dawes 1992). The apex of Vernon's model is a very broad general factor which he calls g, a term first introduced by Spearman (1914). The general factor is subdivided into two broad group
factors: verbal-numerical-educational (v:ed) and practical-mechanical-spatial-physical (km). The broad v:ed factor is subdivided into two minor group factors, verbal and number, and km is subdivided into three minor group factors, mechanical information, spatial, and manual subfactors, which can be subdivided into specific tests. Cattell’s (1971) distinction between crystallized and fluid abilities corresponds roughly to Vernon’s v:ed and km. Crystallized abilities are influenced by education and experience whereas fluid abilities are more basic and more dependent on the functioning of the central nervous system (Bee 1996). Based on the work of Schmid and Leiman (1957), Humphreys (1984) has pointed out that a higher order factor such as g is nothing more than a factor that is defined by more variables than the factors subordinate to it in the hierarchy. Higher order factors are merely broader factors.

More recently, Marshalek, Lohman, and Snow (1983) have proposed a radex model of abilities that described more central, general abilities as residing at the centre of a circle and narrower, more specific abilities around the perimeter. Thus, abilities differed in terms of how central they are (i.e., complexity) and their specific content (i.e., content). More specifically, the model classified tests as complex, intermediate, or simple. Complex tests require abstract problem-solving and rule inferring. More complex tests (e.g., Raven Progressive Matrices, Verbal Analogies) show high correlations with g whereas less complex tests (e.g., Memory Span, Perceptual Speed, Visual Memory) show only small correlations with g. Tests with similar content are labelled and grouped together. Marshalek et al. have pointed out that the radex model is not all that different from the hierarchical model. Tests that measure factors higher in the hierarchical model tend to be more complex, and tests that measure factors lower in the hierarchy tend to be simpler and have lower correlations with g. An advantage of the radex model is that the model more explicitly describes relationships between tests through their placement in the radex. Ackerman (1986, 1988) suggested that a third, speededness dimension be added to Marshalek et al.’s model to address the role of perceptual speed and psychomotor skills, arguing that these skills become more important as the task becomes more speeded.

**Distinctions between Abilities and Related Constructs**

Traditionally, abilities that are less stable have been called skills. Indeed, researchers (e.g., Ackerman 1989a; Henry and Hulin 1989; Murphy 1988) have discussed abilities and skills as if they were dichotomous, maintaining that abilities are more stable than skills. Murphy (1988) further distinguished between abilities and skills, suggesting that skills are learned over a shorter period of time. We agree with Murphy but contend that it is more useful to consider abilities and skills as residing along a continuum of stability with those attributes called abilities residing at the more stable end and those attributes called skills residing at the less stable end of the continuum. Not all abilities are equally stable or all skills equally unstable. Furthermore, a variety of factors can affect stability (Arthur, Bennett, Stanush and McNelly 1998). Thus, it is more useful to explicitly consider the stability or malleability of each ability/skill we assess. However, in the interests of traditional terminology, we will continue to use the terms ability and skill, recognizing that skills are often more recently learned, usually tied to a specific curriculum, and tend to decay more rapidly than abilities.

Similarly, researchers have often distinguished between aptitudes (as being more similar to abilities as traditionally discussed) and achievement (as being more similar to skills). However, the distinction between aptitude and achievement is mainly one of intent or purpose – not test content per se. Further, Lubinski and Dawes (1992) based on the work of Cleary, Humphreys, Kendrick and Wesman (1975) suggest that there are four dimensions appropriate to the description of tests and the behaviour repertoires they sample:

1. Breadth of material sampled.
2. Curriculum represented.
3. Recency of learning sampled.
4. Purpose of assessment (i.e., prediction vs. assessment).

This suggests the following hypothesis regarding the stability of abilities over time: abilities (using the term to encompass also aptitudes, skills, and achievement) that are based on broad samples of behaviour, that are not tied to a specific content domain, and that are based on behavioural repertoires acquired over a long period of time are more stable than abilities that are narrow samples of specialized content.

**Temporal Stability of Cognitive Abilities and Related Constructs**

Research examining the stability of abilities (see Anastasi and Urbina 1997; Bee 1996 for reviews) has shown that general cognitive ability is relatively stable through middle age and only declines in late life. Indeed, longitudinal research (e.g., Eichorn, Hunt and Honzik 1981; Hertzog and Schaie 1986, 1988; Schaie 1994) has suggested that individual differences in g are highly stable over extended periods whereas...
more specific abilities vary in their stability. Performance on intellectual tasks that are speeded, performance-based, or involving fluid intelligence show greater decrements and at younger ages whereas intellectual tasks that are unspeeded, verbal, or involving crystallized intelligence remain stable longer, not demonstrating decrements until the seventies (Bee 1996).

These results can be explained in relation to models of ability. For example, one would expect broad ability factors in the hierarchical model, particularly Vernon’s v:ed, to be more resistant to change due to practice or experience because they encompass more of an individual’s behavioural repertoire that has been accumulated over time. Very specific factors that traditionally have been called skills typically encompass less of an individual’s behavioural repertoire and have been accumulated more recently. Thus, g and the broad group factors should be most stable and the very specific factors more likely to change due to practice or experience. Additionally, recent research has suggested that further research is needed to examine the stability of tests that have been viewed as the most direct measures of g. For example, Flynn (1987) observed that scores on intelligence tests over the past four decades increased more for the Raven Progressive Matrices, compared to multi-content ability measures such as the Wechsler Adult Intelligence Scale. This research suggests that a broad general factor measured by a composite of a wide variety of tasks might be more stable than a factor measured by a more limited sample of abstract problem-solving tasks. Flynn’s hypothesis is that there is no general increase in intelligence but only an increase in fairly narrow problem-solving ability. We should note, though, that tests measuring g are more stable than tests of specific factors.

**Implications of Ability Construct Operationalizations**

The operationalization of abilities can have important effects on observed ability-performance relationships. Operationalizing ability with a measure of g or the broad group factors should result in a predictor that is quite stable, whereas operationalizing ability with a measure of more specific factors is likely to result in a predictor that changes with practice or experience. Thus, it is important to have information about the stability of an ability (e.g., g versus a specific factor) used as a predictor because changes in abilities due to learning, practice, or experience will often result in changes in predictor-criterion relationships. Smith (1994) addressed a similar issue, distinguishing between universals (i.e., characteristics relevant to most jobs such as general cognitive ability), occupationalals (i.e., characteristics that are relevant to particular occupations), and relationals (i.e., characteristics that are relevant to the way a person relates to others in a particular work setting). Smith contended that predictors based on universals (e.g., cognitive ability) remain stable over long periods of time because the nature of work changes slowly whereas predictors based on occupationalals or relationals should be less stable over time.

Additionally, motivational factors can influence the assessment of abilities. As mentioned above, we can assess either typical or maximal performance on predictors and criteria. For example, an individual’s test-taking motivation can affect the extent to which his/her ability is reflected in an ability predictor (Arvey, Strickland, Drauden and Martin 1990). Similarly, Chan and his associates (Chan, Schmitt, Sacco and DeShon 1998; Schmitt, Chan, Sacco, McFarland and Jennings 1999) have found that test reactions and test-taking motivation affect test performance. Indeed, Winterbottom, Pitcher and Miller (1976) noted higher test taking motivation when the LSAT was administered for law school selection decisions than when the LSAT was administered for research purposes after graduation. Interestingly, Ackerman (1994, Ackerman and Heggestad 1997) has noted that traditional ability measures have been designed to assess maximal performance, but predicting typical performance might be more important when predicting long-term academic or occupational performance. We suggest that the typical validation process assesses maximal performance on the ability measure and typical performance in the performance criterion.

Furthermore, we note that motivation plays a role in predictor-criterion relationships through its effects as a moderator. There is a large body of literature demonstrating the direct effects of motivation on performance as a criterion (see Kanfer 1990, for a review). However, Kanfer and Ackerman (1989) have found that motivation and cognitive ability interact in their effects on performance. Their motivational intervention (goals) had stronger effects for low-ability than for high-ability subjects. On the other hand, Sackett, Gruys and Ellingson (1998) have offered strong evidence that two personality variables (need for achievement and dependability) do not interact with cognitive ability but have an additive effect on performance. Additional research is needed to examine potential interactions involving other personality factors (see Barrick and Mount 1991, Digman 1990; Fiske 1949; Goldberg 1993; Hogan 1991; John, Angleitner and Ostendorf 1988; Norman 1963; Tuples and Christal 1961, 1992 for discussions of personality factors and the Big Five model; see Block 1995 for a contrarian view).
In summary, we suggest that it is useful to conceive of abilities as residing on a continuum of stability. Ability measures that broadly sample an individual’s behavioural repertoire and are based on long-term learning (typically referred to as abilities) are hypothesized to be more stable than abilities based on narrow measures of recently learned material (typically referred to as skills). Further, we note that the nature of the measures used (i.e., adequacy of domain sampling) affect (and complicate) the interpretation of ability as a predictor.

Proposition 2: Abilities that are based on broad samples of behaviours will be more stable than abilities based on narrow samples of behaviour. That is, abilities based on broad samples of behaviours will be less susceptible to the effects of training, experience or practice and will demonstrate less change over time, compared to abilities based narrow samples of behaviour.

Task Variables

Researchers have focused on the role of task characteristics, especially task complexity and consistency, in moderating ability-performance relationships (e.g., Ackerman 1986, 1988). Ackerman (1986, 1988) especially has studied the role of task complexity and consistency in performance assessed at different stages of skill acquisition. Other researchers have attempted to extend these results to the job level, suggesting that cognitive will be a stronger predictor of job performance when job tasks are too complex to be automatized (Schmidt et al. 1992). However, much of the research on skill acquisition has been conducted at the task level, and as Schmidt et al. (1992) have pointed out, a significant problem in describing job complexity in terms of job tasks is that many jobs contain tasks that vary widely in complexity so that the job as a whole cannot be simply characterized as simple or complex. Recently, researchers (Quinones, Ford and Teachout 1995; Tesluk and Jacobs 1998) have examined complexity at task and job levels, discussing issues related to skill acquisition at the job or career level in terms of work experience. We further address this issue, describing possible relationships between constructs at the task and job level as well as addressing an additional task variable, task definition/structure, that has received little attention to date in research on dynamic criteria. Thus, in the current section, we examine the role of task variables, specifically task complexity, task consistency, and task definition/structure, in ability-performance relationships and how their effects are influenced by skill acquisition. In the subsequent section, we first posit possible links between task variables and job variables and then examine the role of job variables in ability-performance relationships.

First, we examine the role of task complexity and task consistency in ability-performance relationships at different levels of skill acquisition. Definitions and models of task complexity, consistency, and skill acquisition are followed by a discussion of the role of task definition and task structure. We also note that task variables can have direct effects on performance as well as moderate ability-performance relationships. Finally, we discuss empirical results relating to the role of task characteristics in dynamic criteria.

Task Complexity: Definitions and Models

As Campbell (1988) has noted, there are a variety of ways of conceiving task complexity—as a psychological experience, as a task-person interaction, or as a function of objective task characteristics. Two well-accepted models frame task complexity as a function of objective task characteristics (Campbell 1988, 1991; Wood 1986). Wood (1986) described task complexity in terms of three dimensions: component complexity (number of acts and information cues), coordinative complexity (relationships between acts and information cues), and dynamic complexity (changing relationships between acts and information cues).

Campbell (1988, 1991) offered an alternative perspective, suggesting that any task characteristic that increases information load, information diversity, or rate of information change contributes to task complexity. He identified four task characteristics that met these criteria:

1. the presence of multiple potential ways (i.e., paths) to a desired end-state (i.e., goal);
2. the presence of multiple desired outcomes (i.e., end-states);
3. the presence of conflicting interdependence among paths to multiple outcomes;
4. the presence of uncertain or probabilistic links among paths and outcomes.

Task Consistency: Definition and Model

Schneider and Shiffrin (1977; Shiffrin and Schneider 1977; Schneider, Dumais and Shiffrin 1984) described task consistency as the extent to which the information processing demands of different tasks are consistent or inconsistent. The performance of tasks with consistent information processing demands can be automatized with task practice, freeing attentional resources for other activities (i.e., automatic processing). However, tasks with inconsistent information processing demands require continued high
levels of attention in order to successfully perform the task (i.e., controlled processing). Because task demands are changing, one cannot automate processing of task information and free up attentional resources.

Stages of Skill Acquisition

Models of skill acquisition play a central role in the task complexity and consistency effects observed in dynamic criteria. Specifically, task complexity and consistency effects on ability-performance relationships depend on the stage of skill acquisition. Skill acquisition is generally described as a continuous process with identifiable phases or stages (Anderson 1982, 1983, 1987, 1990; Fitts 1964; Fitts and Posner 1967). For example, Fitts (1964, Fitts and Posner 1967) described three phases of skill acquisition:

1. A cognitive phase, in which the individual learns the task requirements or rules. Memory and reasoning processes are required, so cognitive load is high.


3. A third, autonomous phase, is characterized by very fast performance with very few errors. Performance becomes automatized, and the task requires little conscious attention. Because the cognitive load is low, attention can be devoted to other tasks. Anderson (1982, 1983, 1987, 1990) proposed a similar model of cognitive skill acquisition.

Both Fitts' (1964; Fitts and Posner 1967) and Anderson's (1990) models imply that cognitive ability is less important in the later stages of skill acquisition as performance becomes more automatic. Both models have been tested extensively using a variety of laboratory tasks and thus accurately describe skill acquisition in laboratory tasks.

Interactions between Task Complexity, Consistency, and Skill Acquisition

Ackerman has most clearly addressed the interaction between task context and experience factors. He proposed a model describing the changing effects of abilities (i.e., general cognitive ability, perceptual speed, psychomotor ability) on performance during skill acquisition (Ackerman 1986, 1987, 1988, 1989a, 1989b, 1992). Ackerman hypothesized that for tasks with consistent information processing demands, general cognitive ability will decrease in importance in later trials, perceptual speed will increase then decrease in importance over trials, and psychomotor ability will increase in importance over trials. In contrast, he hypothesized that for tasks with inconsistent processing demands, general cognitive ability will remain an important predictor of performance, and perceptual speed and psychomotor speed will have constant and less important effects on performance. Results using a variety of tasks (e.g., reaction time, air traffic control simulations) and populations (e.g., college students, employees) have generally supported his predictions (e.g., Ackerman 1988, 1992; Ackerman and Kanfer 1993; Kanfer and Ackerman 1989).

For Ackerman, task consistency determines whether general cognitive ability declines in importance with practice or experience as a predictor of performance. Furthermore, he suggested that task complexity affects the speed with which one moves through the stages of skill acquisition. Thus, general cognitive ability should be a more stable predictor of performance for inconsistent and/or complex tasks.

Ackerman (1987, 1988) concluded that distinguishing between tasks of varying consistency and complexity and focusing on different stages of skill acquisition would clarify research on dynamic criteria. His predictions are consistent with the Fleishman–Woodrow changing-task model, suggesting that the relative importance of different abilities changes with varying levels of practice or experience. However, investigators should be cautious in generalizing Ackerman’s results to work performance. He obtained stronger results generally for simpler and more consistent tasks, and other researchers (i.e., Deadrick and Madigan 1990; Pieper 1993; Schmidt, Hunter, Outerbridge and Goff 1988) have not observed similar results in the field using intact jobs. Thus, Ackerman’s results are consistent with prior research on dynamic criteria conducted in laboratory settings. However, the implications of Ackerman’s results for work performance is less clear. Possibly, similar results will be observed for jobs consisting of one or a few tasks.

Implications for Ability-Performance Relationships

Ackerman (e.g., 1987, 1988) has offered substantial research supporting the existence of main and interactive effects for task complexity, task consistency, and skill acquisition on performance and ability-performance relationships. However, additional research is needed to replicate his results in other task or job contexts. Thus, we offer the following propositions to guide future research. The relationships proposed below are displayed in Figure 1.

Proposition 3: Skill acquisition will have a direct effect on criterion task performance and moderate the cognitive ability-performance relationship. That is, one would expect (1)
improvements in task performance; and (2) cognitive ability to decline in importance as a predictor of performance as an individual moves through the stages of skill acquisition.

Proposition 4: Task complexity will have a direct effect on criterion task performance and moderate the cognitive ability-performance relationship. That is, one would expect (1) lower performance on complex tasks; and (2) cognitive ability to be a more important predictor of performance on complex, compared to simple tasks.

Proposition 5: Task consistency will have a direct effect on criterion task performance and moderate the cognitive ability-performance relationship. That is, one would expect (1) lower performance on inconsistent tasks; and (2) cognitive ability to be a more stable and more important predictor of performance on inconsistent, compared to consistent tasks.

Proposition 6: Skill acquisition and task consistency jointly will moderate the cognitive ability-performance relationship. That is, on a consistent task, one would expect cognitive ability to decrease in importance as a predictor of performance as an individual moves through the stages of skill acquisition. However, on an inconsistent task, one would expect cognitive ability to remain a stable predictor of performance across stages of skill acquisition.

Proposition 7: Skill acquisition, task complexity, and task consistency jointly will moderate the cognitive ability-performance relationship. That is, one would expect cognitive ability to be most stable and important in the prediction of performance on complex, inconsistent tasks and to decline in importance as a predictor of performance in simple, consistent tasks as individuals move through stages of skill acquisition.

Proposition 8: Task definition/structure will moderate the cognitive ability-performance relationship. That is, cognitive ability will be a better predictor of criterion task performance when task definition/structure is low.

Task Definition/Structure
Finally, although it has not been addressed in research on dynamic criteria, we posit that task definition/structure plays an important role in ability-performance relationships. Task definition/structure can be understood in terms of models of problem solving (e.g., Getzels 1964; Neisser 1976; Newell and Simon 1972; Reitman 1965; Simon 1973, 1978; Wood 1983). Research (e.g., Arlin 1989) on problem solving has distinguished between problem definition (i.e., whether the individual has identified or discovered a problem) and problem structure (i.e., whether a standard method exists to solve the problem). Furthermore, researchers have proposed typologies of problem definition and/or structure. For example, Wood (1983) and Getzels (1964) have described problem types ranging from those in which the problem is known and methods for solving it are known to the problem solver (i.e., a very well-structured problem) to those in which the problem exists but has yet to be identified and no standard methods for solving the problem are known to the solver or others (i.e., an ill-structured problem).

Researchers also have distinguished well-structured from ill-structured problems, based on whether it is possible to confirm that an optimal or right answer exists (Wood 1983). Some researchers have described ill-structured problems as those in which innovation and creativeness is required for solution (Dillon 1982; Getzels 1964; Mumford, Baughman, Threlfall, Ulman and Costanza 1993; Neisser 1976). Others have suggested that problems in which initial state, goal states and/or operators are unknown are essentially ill-structured (Chi, Glaser and Rees 1982; Reitman 1965). Some researchers (e.g., Simon 1973, 1978, Voss and Post 1988) have maintained that the same mental processes are required for both problem types, suggesting that solving ill-structured problems is a process of identifying and solving well-structured subproblems. However, other researchers (e.g., Neisser 1976) have suggested that different mental processes, those associated with creativity or intelligence, are required for ill-structured problems.

Implications for Ability-Performance Relationships
We posit that task definition/structure moderates ability-performance relationships. That is, tasks that are ill-defined and/or ill-structured will require greater creativity and/or intelligence to solve, compared to well-defined and/or well-structured tasks. Thus, ability-performance validity coefficients should be higher for more ill-defined or ill-structured tasks.

Proposition 8: Task definition/structure will moderate the cognitive ability-performance relationship. That is, cognitive ability will be a better predictor of criterion task performance when task definition/structure is low.

Summary
In our model, we posit that task complexity, task consistency, and task definition/structure moderate ability-performance relationships. More specifically, we conclude that (1) in general the types of tasks assessed in laboratory research on dynamic criteria tend to vary in characteristic
ways on the three dimensions of task complexity, task consistency, and task definition/structure; (2) these task dimensions have been very helpful in interpreting the often mixed results relating to dynamic criteria research in different settings; and (3) examining tasks along these three dimensions will aid in understanding results from future research on dynamic criteria.

**Job Variables**

Researchers have attempted to extend the results based on task level data to the job level. For example, Schmidt et al. (1992) posited that cognitive ability will be a stronger predictor of job performance when job tasks are too complex to be automatized. However, inability to replicate Ackerman’s (e.g., 1987, 1988) results in work settings (i.e., Deadrick and Madigan 1990; Pieper 1993; Schmidt, Hunter, Outerbridge and Goff 1988) and questions about how to define and operationalize job level constructs (e.g., Schmidt et al. 1992) have limited these efforts. As Schmidt et al. (1992) have pointed out, a significant problem in describing job complexity in terms of job tasks is that many jobs contain tasks that vary widely in complexity so that the job as a whole cannot be simply characterized as simple or complex. In this section, we discuss recent models of work experience that directly address different levels of specificity and offer a conceptual basis for hypothesizing about the effects of work experience and job complexity (Quinones et al. 1995; Tesluk and Jacobs 1998). We extend this work by positing additional relationships between task and job level constructs. Thus, in the current section we first identify and define job variables relevant to dynamic criteria, then we posit possible links between task variables and job variables, and finally, examine the role of job variables, i.e., job complexity, job consistency, and task interdependence, in ability-performance relationships. We also address the role of work experience in those effects. As before, posited relationships are shown in Figure 1.

**Job Complexity**

Several researchers have described job complexity in terms of decision making requirements or information processing demands (Gutenberg, Arvey, Osburn and Jeanneret 1983; Hunter and Hunter 1984). However, Murphy (1989b) has noted that jobs in the same complexity category could vary widely in the ‘type, frequency, and intensity of cognitive demands’ of the different tasks performed in the job. We offer an alternate conceptualization of job complexity, drawing from models of task complexity (e.g., Wood 1986). Jobs reflect aggregations of tasks. Hence, jobs could vary in complexity because they are composed of differing numbers of component tasks (i.e., component job complexity) or varying number of relationships between tasks in an individual’s job (i.e., coordinative job complexity). Also, tasks within jobs can vary in complexity. Thus, job complexity can be examined along each of these three dimensions – number of tasks, relationships between tasks, and average task complexity.

**Job Consistency**

Similarly, researchers have begun to address the notion of job consistency. For example, Murphy (1989b) has suggested that the specific ‘tasks, responsibilities, and expectations of individuals working in a particular job may change systematically as a function of job tenure’. He refers to periods in which employees experience high levels of change in their jobs as transition stages and periods of low levels of change as maintenance stages. Transition stages occur when new procedures and/or new tasks have to be learned. When job procedures and/or tasks are mastered, incumbents enter a maintenance stage. Thus, Murphy has suggested that ability-performance relationships are stronger during transition (i.e., learning) stages and that other factors (e.g., motivation, personality) increase in importance as predictors of performance during maintenance stages.

We posit that models of task consistency/inconsistency (Schneider and Shiffrin 1977; Shiffrin and Schneider 1977; Schneider, Dumais and Shiffrin 1984) have implications at the job level. Specifically, jobs can be inconsistent due to changes in the mix of tasks within a job as well as the inconsistency of tasks. Thus, job consistency could result from (a) changes in the mix of tasks that make up a job; and/or (b) the amount of inconsistency within specific job tasks. Transition stages would occur when tasks change and would be longer for jobs comprised of inconsistent tasks.

Research using LSAT scores to predict law school performance provides support for our conclusion that changes in the tasks performed influence predictor-criterion relationships. Several studies have shown that LSAT scores predict first year law school GPAs better than second and third year GPAs (Carlson and Werts 1976; Lin and Humphreys 1977; Powers 1982; Winterbottom et al. 1976). Analysis of the law school curriculum suggests that the curriculum in the second and third years puts more emphasis on skills such as oral presentations that are not measured by the LSAT (Humphreys and Taber 1973).
Work Experience

Finally, in another parallel to a task level construct, i.e., skill acquisition, researchers have addressed the role of cognitive ability in performance at varying levels of work experience. For example, researchers have shown that work experience and cognitive ability both strongly predict job knowledge which in turn is related to performance on work sample tests and supervisor ratings (e.g., Hunter 1983; Schmidt, Hunter and Outerbridge 1986; Schmidt et al. 1992). Further, when work experience is partialled out, cognitive ability assumes a more important role in predicting job knowledge. In addition, Schmidt et al. (1986) have shown that work experience moderates the ability-job knowledge relationship. These findings suggest that variations in work experience can moderate the relationship between general cognitive ability and subsequent outcomes although more research is needed to identify the precise conditions under which this occurs.

Moreover, as Tesluk and Jacobs (1998) and Quinones et al. (1995) have pointed out, work experience is a complex, multidimensional construct. They note that work experience has typically been operationalized as time in job. However, a more complete definition of work experience would address also amount (number of times a task is performed) and type (task experience) would address also amount (number of times a task is performed) and type (task experience) would address also amount (number of times a task is performed) and type (task experience). They note that work experience has typically been operationalized as time in job. However, a more complete definition of work experience would address also amount (number of times a task is performed) and type (task experience) would address also amount (number of times a task is performed) and type (task experience) would address also amount (number of times a task is performed) and type (task experience). They note that work experience has typically been operationalized as time in job. However, a more complete definition of work experience would address also amount (number of times a task is performed) and type (task experience).

Implications for Ability-Performance Relationships

We suggest that complexity, consistency, and learning-related constructs at the task level have parallel effects at the job level. Just as task complexity, task consistency, and skill acquisition have main and interactive effects on criterion performance and ability-performance relationships, similar effects can be examined at the job level. (We note, though, that job constructs tend to be multidimensional, whereas task level constructs tend to be unidimensional.) Thus, our propositions are derived from parallel research at the task level (e.g., Ackerman 1987, 1988) as well as recent research offering initial support at the job level (e.g., Murphy 1989b; Schmidt et al. 1986). For example, Murphy’s and Schmidt et al.’s research has suggested that work experience moderates ability-performance relationships. Posited relationships are stated below and shown in Figure 1.

Proposition 9: Work experience will have a direct effect on criterion task performance and moderate the cognitive ability-performance relationship. That is, one would expect (1) improvements in job performance; and (2) cognitive ability to decline in importance as a predictor of job performance as an individual obtains increased levels of work experience.

Proposition 10: Job complexity will have a direct effect on criterion job performance and moderate the cognitive ability-performance relationship. That is, one would expect (1) lower job performance; and (2) cognitive ability to be a more important predictor of job performance on complex jobs.

Proposition 11: Job consistency will have a direct effect on criterion job performance and moderate the cognitive ability-performance relationship. That is, one would expect (1) lower job performance; and (2) cognitive ability to be a more stable and more important predictor of job performance on inconsistent jobs.

Proposition 12: Work experience and job consistency jointly will moderate the cognitive ability-performance relationship. That is, on a consistent job, one would expect cognitive ability to decrease in importance as a predictor of job performance as an individual obtains higher levels of work experience. However, on an inconsistent job, one would expect cognitive ability to remain stable in its prediction of job performance at varying levels of work experience.

Proposition 13: Work experience, job complexity, and job consistency jointly will moderate the cognitive ability-performance relationship. That is, one would expect cognitive ability to remain stable and important in the prediction of job performance on complex, inconsistent jobs and to decline in importance in simple, consistent jobs as individuals obtain higher levels of work experience.

Task Interdependence

Finally, we posit that one additional job level variable, i.e., task interdependence, will play a role in cognitive ability-performance relationships. Task interdependence is defined in terms of interactions between people as well as a characteristic of tasks. Research has shown that task interdependence affects variables important to performance such as motivation (e.g., Hirst 1988; Kiggundu 1983) and group effectiveness (e.g., Saavedra, Earley and Van Dyne 1993). Some researchers (e.g., Kiggundu 1981; Saavedra...
et al. 1993; Salas, Dickinson, Converse and Tannenbaum 1992; Shea and Guzzo 1987) described task interdependence in terms of the degree to which the interaction among work group members affects performance of the task. These social interaction definitions of interdependence tend to be broad, often including task, feedback, goal, and/or resource (e.g., information, materials) interdependence (Saavedra et al. 1993; Wageman 1995). Other researchers have described task interdependence more narrowly, in terms of whether tasks can be divided into independent parts (Steiner 1972; Thompson 1967).

Implications for Ability-Performance Relationships

We posit that task interdependence moderates ability-performance relationships. When tasks within an individual’s job are interdependent with tasks performed by another individual, it is more difficult to identify any individual’s unique contribution to performance and less likely that performance measures will accurately reflect any individual’s ability (Jones 1984). Thus, ability-performance validity coefficients should be lower for jobs with interdependent tasks.

Proposition 14: Task interdependence will moderate the ability-performance relationship. That is, cognitive ability will diminish in importance when task interdependence is high.

Summary

We suggest that research and models on task level variables offer guidance in defining job level variables likely to be relevant in research on dynamic criteria. Integrating previous models with our extensions provides researchers with a framework for relating task and job level variables, including variables related to complexity, consistency, and learning (i.e., skill acquisition and work experience). Additionally, we posit that job complexity, job consistency, and task interdependence moderate ability-performance relationships and that these affects can be influenced by work experience. More specifically, we conclude that examining jobs along these dimensions and understanding links between task and job level variables will increase our understanding of processes involved in dynamic criteria.

Organizational Variables

Finally, we propose that four organizational variables are important in understanding dynamic criteria: technology changes, changes in work processes, situational constraints, and situational strength. Additionally, we discuss an interactional perspective for understanding dynamic criteria. We posit that organizational variables moderate cognitive ability-performance relationships and in some instances have a direct effect on criterion performance. We discuss the effects of organizational variables in the context of two mechanisms. The first mechanism relates to the role of organizational variables in the context of two mechanisms. The second mechanism relates to the role of organizational constraints, such as situational constraints or situational strength in ability-performance relationships.

We suggest that contextual variables at the level of the organization that are unrelated to individuals’ abilities or tasks could affect criterion performance or ability-performance relationships. These have been referred to by Campbell and Pritchard (1976, p. 65) as facilitating and inhibiting conditions not under the control of the individual. Although these factors have received attention in the areas of personality (Barrick and Mount 1993; Frink and Ferris, in press; Robie and Ryan, in press) and human factors (see Chapanis 1976; Howell 1991 for reviews), contextual factors in the organization have received little research attention in I/O psychology.

Changes in Technology and Changes in Work Processes

We posit that organizational events play a role in ability-performance relationships. For example, researchers have suggested that changes in technology or changes in work processes could trigger additional transition stages (Murphy 1989b, see above). For the purposes of this article we define changes in technology as changes made to the way work is done resulting from changes in available technology. We define changes in work processes as changes made to how the work is done primarily resulting from organizational decisions, for example, a strategic decision to shift attention from one product to another. The distinction between changes in technology versus work processes is consistent with a sociotechnical perspective that focuses on both technical and human/processual issues in organizational functioning.

Proposition 15: Changes in technology or work processes will have a direct effect on criterion job performance and moderate the cognitive ability-performance relationship. That is, one would expect (1) lower job performance; and (2) cognitive ability to
remain a more stable and important predictor of job performance when jobs are changing due to changes in technology or work processes.

Situational Constraints and Situational Strength

Additionally, we posit that some organizational variables, such as situational constraints or situational strength (for example, due to high formalization or a strong climate), can result in decreased ability-performance relationships (Schneider 1978; James, Demaree, Mulaik and Ladd 1992). Peters and O’Connor (1980, 1988) argued that situational constraints (1) create a ceiling effect in performance for high ability individuals; (2) limit the range and variance of observed performance (3) attenuate correlations between individual difference factors (e.g., ability, motivation) and performance; and (4) lead to dissatisfaction and frustration. Laboratory research has supported their contentions, indicating that various situational constraints lead to decreased performance and increased dissatisfaction and frustration (Peters, Chassie, Lindholm, O’Connor and Kline 1982; Peters, Fisher and O’Connor 1982; Peters, O’Connor and Rudolf 1980).

Situational strength (e.g., Adler and Weiss 1988) might play a similar role in ability-performance relationships. Mischel (1977a, p. 347) explained the concept of strong situations:

> Psychological ‘situations’ (stimuli, treatment) are powerful to the degree that they lead everyone to construe particular events the same way, induce uniform expectations regarding the most appropriate response pattern, provide adequate incentives for performance of that response pattern, and require skills that everyone has to the same extent.

Hesketh and Robertson (1993), in their conceptual model of dynamic criteria, have suggested that one would expect person variables (ability, motivation, or personality) to account for performance in weak situations, whereas environmental or situational factors would have more effect in strong situations.

Thus, research suggests that some factors in the organization, such as situational constraints or situational strength, function to increase uniformity of employees’ behaviour. Under these uniformity-creating conditions, one would expect the employee to have less opportunity to demonstrate his ability.

Proposition 16: Situational constraints or situational strength will moderate the cognitive ability-performance relationship. That is, one would expect cognitive ability to decrease in importance as a predictor of criterion performance under high levels of situational constraints or situational strength.

An Interactionist Perspective

The interactionist perspective developed in the 1960s (e.g., Endler and Magnusson 1976b) helps to clarify our understanding of the role of contextual factors in ability-performance relationships. Schneider (1983) suggested that interactional psychology offered a solution to the debate between trait theorists’ claims regarding the primacy of internal characteristics (e.g., Allport 1966) and the situationists’ claims regarding the primacy of external factors (e.g., Mischel 1968) as determinants of behaviour. Bowers (1973) and other researchers (e.g., Endler and Magnusson 1976a; Magnusson and Endler 1977; Mischel 1977b; Pervin and Lewin 1978) offered compelling arguments that, in fact, persons and situations interacted in their effects.

There have been two major conceptions of interactions in interactional psychology. First, a mechanistic view defines interactions in terms of person by situation interactions such as can be obtained by partitioning variance into main effect and interaction components (Endler and Magnusson 1976a; Magnusson and Endler 1977). Second, an organismic view defines interactions in terms of reciprocal causation, that is, persons affect their situations and situations affect persons (Endler and Magnusson 1976a; Magnusson and Endler 1977). Whereas the mechanistic view conceives the person as passive, the organismic view conceives the person as an active agent, influencing as well as being influenced by situations over a period of time. Many researchers (Endler 1988; Endler and Magnusson 1976a; Fiske 1977; Magnusson and Endler 1977; Raush 1977; Wachtel 1977) have argued that reciprocal causation, i.e., continuous, multidirectional effects over time, should be the primary focus of interactional psychology.

We believe that reciprocal causation is especially important in situations where the individual can exert substantial control over the situation and the tasks performed. Studies of undergraduate college performance in which the individual can to a considerable extent pick and choose majors, courses, instructors, etc. is a good example of the possible potent effects of reciprocal causation. Reciprocal causation suggests that methodological improvements can be achieved by studying the larger context in which the individual is making choices. In the example of college studies, we need more information than just GPAs. We need to collect information about individual changes in majors, courses, etc. in order to understand the extent to

© Blackwell Publishers Ltd 2000
which reciprocal causation explains the decline in validity coefficients in predicting undergraduate grade point average as compared with other hypotheses such as the changing-subject or changing-task models.

Although interactional psychology has been prominent in personality (e.g., Magnusson and Endler 1977) and developmental psychology (e.g., Lerner 1985a, 1985b), some researchers have applied interactional psychology to I/O topics, including interviewing (Eder and Buckley 1988), perceived equity (Schneider 1978), climate (e.g., Bray, Campbell and Grant 1974), and job performance (Campbell et al. 1993; McCloy, Campbell and Cudeck 1994). Finally, most directly relevant to dynamic criteria, Hanges et al. (1990) gave a clear demonstration of how person factors interact with situational factors to affect stability of performance over time. Hofmann and his associates (Hofmann et al. 1992; Hofmann et al. 1993), drawing from an interactional developmental theory proposed by Lerner (1985a, 1985b), arrived at a similar conclusion, specifically that ‘individual and environmental opportunities and constraints interact over time, each influencing the other’. The individual can change his/her environment in ways that enhance or detract from his/her performance while in turn the environment changes the individual.

Research on interactional psychology has raised two important methodological issues: (1) the need for a taxonomy of situations (Edwards 1991; Frederiks 1976; Hattrup and Jackson 1996; Hesketh 1993; Magnusson 1976, 1981; Mischel 1977a, 1977b; Murtha, Kanfer and Ackerman 1996; Sells 1976) and (2) the need for better analysis methods, including causal modeling (Murtha et al. 1996), Markov chains or path analyses (Endler 1988), response surface methodology (Edwards 1991), and profile comparisons (Caldwell and O’Reilly 1990).

Implications for Ability-Performance Relationships

Interactional psychology and reciprocal causation provide important models for exploring the role of contextual factors in dynamic criteria. We note that the individual can influence his/her environment by changing (1) tasks within the current job; (2) jobs within the current organization; or (3) organizations. However, we suggest that how individuals influence changes in tasks within their current job is the most relevant issue for dynamic criteria.

Proposition 17: Under some conditions, an individual can change his/her environment, for example through job variables, in ways that enhance or detract from his/her performance while in turn the individual is changed by the environment. Thus, the individual and the environment have reciprocal effects.

Summary

We suggest that organizational variables can influence ability-performance relations through two different mechanisms. Some variables such as changes in technology or work processes can reduce consistency in the job and thus increase the importance of ability as a predictor of performance. Other variables such as situational constraints or situational strength function to increase uniformity in performance and so decrease the importance of ability as a predictor of performance. Finally, we suggest that evidence from interactional psychology suggests that a complete model of dynamic criteria needs to include the notion that individuals influence as well as are influenced by their environments.

Implications for Future Research

On the basis of our review and discussion of dynamic criteria research, we propose that further progress will be facilitated by addressing a number of important conceptual and methodological issues: the definition of dynamic criteria, measurement of predictors and criteria, the concept of reciprocal causation, appropriate statistical techniques, problems with research samples, and factors affecting generalizability.

Clarifying the Definition

Barrett et al.’s (1985) review of dynamic criteria research distinguished between three definitions for the construct: (1) changes in average group performance; (2) changes in the rank order of subjects on repeated performance measures; and (3) changes in predictive validity coefficients. Barrett et al.’s (1985) second definition has tended to focus research on changes in performance over time, and the effects of these changes on the validity of predictor variables. We believe that research on dynamic criteria should focus on systematic changes (either increases or decreases) in predictive validity coefficients, and the many variables that may influence these changes.

Measuring Predictors

We posited in our model that abilities differ in their stability. However, researchers (e.g., Fleishman and Hempel 1954, 1955, Humphreys 1968) typically have measured ability only once, precluding an assessment of the stability of the measured ability. Multiple measures of ability are
needed to determine whether subjects’ abilities are changing. Additionally, care should be taken concerning the content of the ability measure. Although some researchers (Alvares and Hulin 1972; Henry and Hulin 1987, 1989) have argued that early performance on a task and performance on an ability test are manifestations of the same underlying phenomenon, other researchers (Ackerman 1989b) have properly criticized this approach. We suggest that performance typically is a function of several underlying constructs—not just ability.

Measuring the Criterion

Researchers frequently have focused on the simplex-like pattern of correlations among repeated measures of performance as evidence of dynamic criteria (e.g., Humphreys 1968). However, whereas the simplex-like pattern is ubiquitous, it does not necessarily signal a decrease in validity coefficients (Ackerman 1989b; Deadrick and Madigan 1990; McEvoy and Beatty 1989) or even a change in validity coefficients (Schmidt 1993). Thus, research on dynamic criteria should focus on changes in predictor-criteria relationships and the many variables that affect these changes.

Reciprocal Causation

The concept of reciprocal causation offers a methodological improvement to dynamic criteria research, particularly when considering the role of contextual factors. We tend to think of the effects of the context on the individual, e.g., constraining or providing opportunities for increasing skills or abilities. However, it is important to note that the individual also can exert control over the situation and the tasks performed, e.g., seeking tasks that match his/her abilities. Reciprocal causation, how the individual influences and is influenced by his/her environment, has a strong effect on predictor-criterion relationships.

Statistical Techniques

Some statistical techniques used in dynamic criteria research have not been entirely appropriate. For example, Fleishman (1953; Fleishman and Hempel 1954, 1955) and others factor analysed predictors along with repeated measures of criterion performance. Other researchers (Bechtold 1962; Humphreys 1960) have correctly pointed out that exploratory factor analyses are inappropriate for simplex patterns of correlations and for data that are not experimentally independent. More recent research (Fleishman and Mumford 1989) has factor analysed the test battery separately and then projected the loadings of trial stages on factors using Dwyer’s extension (Dwyer 1937).

Another issue is analysing patterns of correlations versus testing the significance of adjacent pairs of correlations. Barrett et al. (1989) tested differences between pairs of correlations to determine whether there was evidence of dynamic criteria. However, we believe the real issue when considering evidence of dynamic criteria is whether there is systematic change in the pattern of correlations—not differences between adjacent correlations.

Additionally, structural equation modelling and multilevel modelling has increased our understanding of dynamic criteria. For example, Joreskog and Sorbom’s (1989) re-analysis of Humphreys’ (1968) data revealed that ACT scores and high school rank had a direct effect on first semester GPAs but only an indirect effect on GPAs in subsequent semesters. Structural equation models have also been used to model changing effects of abilities on performance over time (Fleishman and Mumford 1989) and to study components of performance (Campbell et al. 1993). Alternately, Hofmann, Deadrick and their associates have advocated the study of intra-individual change, that is, examining inter-individual differences in patterns of change within individuals (Deadrick, Bennett and Russell 1997; Hofmann et al. 1993; Hofmann et al. 1992). Using multilevel modelling Hofmann and Deadrick have uncovered systematic differences in the way individuals change with practice or experience. Even more interesting, Deadrick et al. (1997) have found that variables that predicted initial performance (i.e., psychomotor skills) were different from the variables that predicted performance improvement (i.e., cognitive ability). Similarly, Schmitt and Chan (1998) discuss the importance of using complex modeling methods such as causal or multi-level modeling for the study of dynamic criteria. Additionally, Schmitt et al. (1999) offer an example of how item response theory might be used to improve the assessment of traits, such as cognitive ability.

Finally, research on interactional psychology has suggested a number of methods that might be used to better examine reciprocal causation, such as structural equations modelling, Markov chains, response surface methodology, and profile comparisons. Similarly, we suggest that research should examine the usefulness of alternative methods for studying the role of reciprocal causation in dynamic criteria.

Research Samples

A persistent methodological problem in dynamic criteria studies is the attrition of subjects in repeated measurements of performance over
time. Often the result is restriction of range on the performance measure which complicates interpretation of the results. For example, Mauger and Kolmodin (1975) observed correlations between cumulative GPAs and SAT-V and SAT-M of .52 and .43, respectively, in a sample of 318 students who terminated their college work before graduation. In contrast, the standard deviation for 520 graduating seniors was smaller and so were the correlations (.26 and .22). Similarly, Hanges et al. (1990) in their study of repeated measures of student ratings of faculty teaching effectiveness analysed radically different faculty sample sizes ranging from 231 to 79. We suggest that although attrition is unavoidable, presenting constant N data will aid in examining dynamic criteria even though sample sizes may be smaller.

**Generalizability**

Finally, Barrett et al. (1989) argued that it is a bad assumption that laboratory tasks would generalize to ‘real world’ tasks. Similarly, Schmidt et al. (1988) criticized previous research for generalizing inappropriately from narrow and automatable tasks to broader, more complex, and less automatable real world job components. Ackerman (1987, 1988) eloquently argued that ability-performance relationships do change over time – either increasing or decreasing in strength, but much of the confusion in previous research is due to aggregating across tasks of varying consistency and complexity, different stages of skill acquisition, and abilities with different characteristics. We suggest that examining the factors identified in our model will aid in identifying boundaries to claims of generalizability.

**Implications for Professional Practice**

Dynamic criteria impact professional practice in two ways: (1) in assessing the utility of predictors across time; and (2) in evaluating the need to revalidate a predictor battery. The main concern is the stability of general cognitive ability or g since general cognitive ability is the primary cognitive predictor of job performance (Thorndike 1986). We suggest that cognitive ability tests that sample a wide range of an individual’s behavioural repertoire will be more stable than more narrowly focused measures of g such as tests of abstract reasoning. Further, when narrower or more recently acquired skills are used as predictors, the practitioner needs to know over what time period that skill is likely to be stable. Another important issue that needs to be addressed is the nature of the tasks performed by the incumbent. There is evidence that the validity of general cognitive ability over the long term is moderated by task consistency complexity. Also, the extent to which job tasks can be expected to change over time due to reorganizations, re-engineering, or changes in technology can affect the validity of cognitive tests. Practitioners should carefully consider likely changes in valued aspects of job performance. For example, if criteria shift from task proficiency to contextual behaviours, the validity of general cognitive ability may be expected to diminish. Finally, the extent to which an incumbent can shape the job to his/her abilities should be evaluated. The amount of control the individual has over his/her immediate environment seems to be a potent moderator of the validity of general cognitive ability.

**Conclusion**

Whereas recent models have offered more comprehensive and more adequate explanations of dynamic criteria (e.g., Hesketh and Robertson 1993; Murphy 1989b), we believe a more fine-grained model focusing on the ability-performance relationship is needed. To a modest extent, we have shown how task variables, job variables, and organizational variables affect performance and ability-performance relationships and the role of learning-related variables in those effects. However, much work remains to be done. Thus, the task for researchers is to identify which of these factors are relevant in different research situations and to further clarify their effect on the relationships between ability and performance. In conclusion, Ackerman (1987) commented that the simplex pattern, while ubiquitous, is of little help in understanding dynamic criteria – we agree. Analogously, we suggest that debates about whether criteria are dynamic or stable are of little help in increasing our understanding of ability-performance relationships or the nature of human performance. Clearly, cognitive ability is an effective predictor of job performance for a wide variety of jobs (e.g., Hunter and Hunter 1984; Schmidt et al. 1992). The question becomes: what variables affect the validity of cognitive ability not only across jobs but also across time and situations? If our goal is to improve the quality of prediction, we must identify and model variables likely to have systematic effects on ability-performance relationships.

**References**

Ackerman, P.L. (1986) Individual differences in information processing: An investigation of
intellectual abilities and task performance during practice. *Intelligence, 10*, 109–139.


Ackerman, P.L. and Heggestad, E.D. (1997) Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219–245.


Consulting Psychologists Press, Palo Alto, California.


