

**Defining and Measuring Individual Level Job Performance:  
A Review and Integration**

**Maria Rotundo, Ph.D.**

**Joseph L. Rotman School of Management**

**University of Toronto**

**September 2002**

## **Introduction**

To grow and thrive in today's competitive environment, the government, educational institutions, and businesses need to continue to provide value. Institutions and businesses invest a great deal of resources in employment tests, job interviews, and training programs in an attempt to increase motivation, build human capital, and ultimately improve the job performance of employees. The measurement process of individual level job performance consists of many parts including defining and measuring job performance. The last two decades of research have seen progress in defining job performance and in understanding the performance measurement process (Arvey & Murphy, 1998; Campbell, 1990). This literature review covers two areas of the performance measurement process: 1) Theory of job performance; 2) Measuring job performance, which incorporates 360-degree feedback and the stability of job performance.

Despite the obvious importance of understanding individual level job performance, it was only during the last 20 years that researchers took an empirical interest in trying to define this construct. Research articles and book chapters emerged that present different models and approaches to defining job performance (Campbell, 1990; Borman & Motowidlo, 1993; Murphy, 1989). This literature includes advantages and disadvantages of micro definitions that focus on specific individual behaviors versus macro definitions that focus on productivity or effectiveness. Furthermore, some of these ideas were modeled around jobs in the U.S. Army and Navy, while others took a more inclusive approach (Campbell, McHenry, & Wise, 1990; Murphy, 1989).

The first part of the literature review presents an integration and summary of the body of literature that has emerged during the last twenty years in the area of defining job performance. Generally, researchers agree that job performance can be defined on a micro level as actions and

behaviors of an employee that contribute to the goals of the organization (Campbell, 1990; Murphy, 1989). A wealth of literature speaks to issues surrounding the structure underlying job performance (i.e., those categories of behaviors that are valued by supervisors). A review of this literature indicates that for some time the only behaviors that received attention in the research literature and by organizations were those behaviors associated with the production of a good or the provision of a service, namely task performance. Some researchers propose that employees do not spend all of their time at work performing activities that are strictly related to task performance (e.g., Borman & Motowidlo, 1993; Katz & Kahn, 1978; Organ, 1988). For example, employees help coworkers or volunteer to engage in activities that benefit the organization. These behaviors also contribute to the organization in positive ways and have been given a variety of names (e.g., organizational citizenship behavior, contextual performance, extra-role behaviors). Furthermore, other researchers remind us that some employees engage in behaviors that are counterproductive to the organization or its members (Crino, 1994; Giacalone & Greenberg, 1997). For example, employees may fail to follow rules and procedures, steal merchandise, or sabotage company property. In contrast to task or organizational citizenship behaviors, these behaviors detract from the goals of the organization. A review of different models of job performance reveals that these are three distinct groups of behaviors that comprise the domain of job performance.

The most common way to measure job performance is a supervisor or manager's rating of an employee's (ratee's) job performance. Research has shown that supervisors incorporate a great deal of information in these ratings. Performance measurement systems typically focus on the supervisor or manager as the rater. However, a number of corporations use 360-degree performance measurement systems. In these systems performance appraisals of the target

employee are provided by peers, subordinates, and/or the ratee him/herself and are included in the performance appraisal process. Some organizations use the feedback from peers, subordinates, or the self for developmental purposes only, while other organizations incorporate this feedback when making personnel decisions (e.g., merit increase, promotion). However, questions arise as to whether peers, managers, the self, and subordinates value the same aspects of performance when rating others. Research findings are mixed. Although some research indicates that self-ratings of job performance are congruent with supervisory ratings of job performance (e.g., Farh, Werbel, & Bedeian, 1988), other research indicates the opposite (McEnery & Blanchard, 1999). Furthermore, research findings are also mixed as to whether 360-degree feedback improves managerial effectiveness (e.g., Atwater, Ostroff, Yammarino, & Fleenor, 1996; Smither, London, Vasilopoulos, Reilly, Millsap, & Salvemini, 1997).

Another issue that is central to our understanding of the performance measurement process is whether or not an individual's job performance is static or dynamic. Research has begun to investigate this issue and reports mixed results. Evidence supports both a stable and a dynamic construct (e.g., Austin, Humphreys, & Hulin, 1989; Barrett & Alexander, 1989). Hence, future research needs to explore measurement issues that may explain the inconsistency in the findings. Furthermore, future research needs to examine the antecedents and consequences of performance stability in an effort to determine the extent to which job performance is stable or dynamic.

# **Theory of Job Performance**

## ***Job Performance Defined***

Job performance is arguably one of the more important dependent variables of interest to educators, businesses, the government, and society. Researchers and businesses are just now reaching consensus on common definitions and conceptualizations of individual level job performance. Rotundo (2000) presents an integration and summary of the body of literature that has emerged in defining job performance. Key themes from this review are highlighted below combined with a discussion of more recent literature. The end product is a definition of job performance that incorporates the ideas from prior research.

Rotundo (2000) explains that although researchers provide their own conceptualization of job performance, a typical definition focuses on behaviors or actions of individuals, not results or outcomes of these actions and behaviors. Smith (1976) discusses some of the problems with various definitions of job performance and stresses that an accurate measure of job performance includes the direct observation of behavior. Murphy (1989) states that job performance should be defined in terms of behaviors rather than results. He explains that results-based measures are not always functional to the organization, as employees may try to maximize results at the expense of other things. In the job of a retail sales clerk, examples of behaviors include greeting customers, answering questions about store products, demonstrating knowledge of store policy and procedures, where an example of a results measure includes total sales per week or month. In a situation where employees are evaluated on sales only, employees may compromise certain behaviors (being polite to customers) in order to try to maximize sales (e.g., forcing products on customers). Murphy also defines performance as behaviors that are related to the goals of the organization. Campbell (1990) defines performance as those actions or behaviors under the

control of the individual, that contribute to the organization's goals, and that can be measured according to the individual's level of proficiency, a definition that is consistent with the others.

The definitions of job performance reviewed here have some common features. These features include a focus on behaviors that are under the control of individual not results and on behaviors that contribute to the goals of the organization.

Job performance needs to be differentiated from other measures of performance that are sometimes used interchangeably, often incorrectly so. Unlike performance, which focuses on actions and behaviors, Campbell (1990) defines individual effectiveness as *an evaluation of the results* of an action. For example, evaluating the total sales of a retail sales clerk at the end of the day or month against the standards of performance for sales clerks set by the store is an example of a measure of effectiveness, where a measure that rates the sales clerk on appropriate work behaviors is a measure of job performance. Although effectiveness is sometimes used as an index of job performance, it actually measures something different (Campbell, 1990).

Productivity is another term that is often confused with job performance. Productivity has been defined as the ratio of outputs relative to inputs into some production process (Mahoney, 1988). Outputs can include the number of units produced, the quality of the units produced, or the number of units sold, where inputs can include raw materials, time, or effort. Although productivity has been used as an index of how well an organization is doing, Mahoney cautions against its use as an index of job performance because it reflects a different construct.

The purpose of this section was to review how job performance has been defined in research. Researchers strongly recommend that performance be defined in terms of behaviors that are under the control of the individuals and that contribute to the goals of the organization (Campbell, 1990; Murphy, 1989; Smith, 1976). Although it is common to find the words job

performance, effectiveness, or productivity used interchangeably, this practice is inappropriate as these terms reflect different aspects of employees or the organization.

### ***Structure Underlying Job Performance***

Job performance was defined as actions that contribute to organizational goals and that are under the individual's control. Rotundo (2000) argues that this definition includes a wide range of job behaviors and that some behaviors contribute to the employee's duties and responsibilities, while other behaviors still affect the goals of the organization but do not fall under duties and responsibilities. The purpose of this section is to review the literature that attempts to delineate the structure underlying these job behaviors. The goal is to devise a category scheme for classifying job behaviors. Researchers have attempted to classify the behaviors into different components of job performance (e.g., Campbell, McCloy, Oppler, & Sager, 1993; Campbell, McHenry, & Wise, 1990; Katz and Kahn, 1978). Rotundo provides an in-depth review of this literature. This section will present highlights of this review and incorporate the most recent literature.

#### I) Taxonomies of Job Performance

Campbell, McHenry, and Wise (1990) modeled job performance in a set of 19 entry-level Army jobs (i.e., Military Occupational Specialties) and found support for five performance factors. These factors include actions and behaviors that relate to completing tasks, working with others, and maintaining personal discipline. Together, these factors represent a broad range of behaviors that contribute to the goals of the military. Although these components were derived from entry-level Army jobs, they are likely to generalize to jobs in other fields as well.

In an effort to model job performance for all jobs in the *Dictionary of Occupational Titles*, Campbell (1990) proposes an eight-factor model of performance. Rotundo and Sackett

(2002) summarized this information in tables (See Table 1 adapted from Rotundo and Sackett, 2002). Unlike the model proposed by Campbell, McHenry, and Wise (1990), which was intended to model the performance of entry-level Army jobs, this model is intended to be more comprehensive and inclusive of all jobs. Rotundo (2000) notes that five of these factors overlap with the ones proposed by Campbell et al. (1990). The areas of overlap include core task proficiency, non-job specific task proficiency, maintaining personal discipline, demonstrated effort, and facilitating peer and team performance. However, Rotundo notes that maintaining personal discipline is more comprehensive in this model as it includes a wider range of behaviors that are counterproductive in nature. Furthermore, she notes that three components are unique to the eight-factor model. These include written and oral communication, supervision/ leadership, and management /administration. A review of these two taxonomies indicates that task behaviors are included as well as behaviors that involve relating to others and demonstrating personal discipline.

Murphy (1989) proposed a four-category scheme to model a large group of jobs in the Navy (See Table 1). These categories were derived from a set of organizational goals in the Navy. Rotundo (2000) explains that task performance as defined by Murphy is similar to Campbell's Core technical proficiency and Job-specific task proficiency in that all three incorporate task behaviors. Similarly, *Destructive/hazardous behaviors* and *Down-time behaviors* reflect the negative pole of Campbell's *Personal discipline*. These two components of Murphy represent behaviors that deter the individual from accomplishing tasks or prevent the organization from achieving its goals. Murphy's *Leadership* and *Facilitating peer and team performance* include positive interpersonal behaviors.



Although the taxonomies reviewed so far propose different models of job performance, there are groups of behaviors that are common to all models. These behaviors include task or the accomplishment of duties and responsibilities, cooperation or interpersonal behaviors, and deviant acts (Rotundo, 2000).

The previous discussion focused on jobs in general or jobs in the Army or Navy. Borman and Brush (1993) modeled the job performance of managers. More specifically, they proposed a taxonomy of 18 managerial performance requirements, which they further grouped into four categories (See Table 1). These four categories also represent task, interpersonal, and deviant behaviors.

Hunt (1996) analyzed supervisory ratings of non-task elements in a variety of hourly, entry-level jobs. He chose to focus on non-task elements because of the perception that hourly, entry-level jobs require a low level of job-specific knowledge, skills, and abilities. Hunt defines generic work behaviors as “behaviors that influence the performance of virtually any job” (Hunt, p. 53). Exploratory and confirmatory factor analyses of supervisory ratings revealed nine categories (See Table 1). Although these dimensions do not include task behaviors, they do represent the interpersonal and deviant behaviors, which are consistent with the other models of job performance.

The first step in understanding the performance measurement process is to define job performance; the second step is to determine the structure underlying job performance. This section reviewed five different taxonomies of job performance. A clear picture emerges from this review. Job performance can be described by three broad categories of behaviors, which are subsequently labeled task, organizational citizenship, and counterproductive performance (Rotundo, 2000). The first category reflects behaviors that are consistent with performing duties

and responsibilities. The second domain, organizational citizenship, includes behaviors that are clearly related to organizational goals in a positive way but do not necessarily contribute to the core functioning of the organization (e.g., exerting effort, maintaining professional relationships, and supporting and helping others). The third category or domain is counterproductive behavior. It represents negative behaviors that can harm the well-being of the organization or coworkers (e.g., substance abuse, absenteeism, tardiness, theft). The next section reviews research that delves more deeply into each of these components in an attempt to provide a more refined definition of each component and to devise a list of behaviors that comprise each performance component.

## II) Task Performance

According to Borman and Motowidlo (1993) task performance is “the proficiency with which job incumbents perform activities that are formally recognized as part of their jobs; activities that contribute to the organization’s technical core either directly by implementing a part of its technological process, or indirectly by providing it with needed materials or services” (Borman & Motowidlo, p. 73). These researchers define technical core as the set of activities and processes that are used to convert raw materials (e.g., manufacturing) into products the organization produces (See Table 2 adapted from Rotundo and Sackett, 2002). Some jobs contribute directly to an organization’s technical core by way of transforming raw materials, where other jobs contribute indirectly by replenishing raw materials, distributing finished products, or providing support services (e.g., managers, accountants). Borman and Motowidlo believe that behaviors or activities that contribute to the technical core, directly or indirectly, should be included in task performance. Two central features in this definition of task

performance are activities that are formally recognized as part of the job and contribute to the technical core.

Similarly, Murphy (1989) defines task performance as the accomplishment of tasks within an incumbent's job description. Campbell et al. (1993) and Campbell (1990) also include elements related to task performance in their taxonomies of job performance (e.g., core technical proficiency, general soldiering proficiency, job-specific task proficiency, and non-job-specific task proficiency).

Researchers conceptualize task performance as behaviors that contribute directly or indirectly to the technical core and behaviors that are recognized as part of the job or job description. However, Rotundo (2000) notes that restricting a definition of task performance (or any aspect of job performance) to include only those behaviors listed in a job description is problematic because job descriptions for the same job may differ from one organization to the next, which makes it difficult to compare performance across organizations. Furthermore, Rotundo notes that jobs are constantly changing without these changes being reflected in job descriptions. Therefore, measures of performance that depend on the content of a job description may not be accurate. On the basis of these arguments, Rotundo recommends that task performance be defined as actions and behaviors that contribute directly or indirectly to the production of a good or the provision of a service.

### III) Organizational Citizenship Performance

Traditionally, organizations and researchers focused on task performance. However, since the 1980's, researchers believe that there are additional activities that are relevant in other ways to the goals of the organization (Borman & Motowidlo, 1993; George & Brief, 1992;

Organ, 1997). This section will review the literature on what will be subsequently referred to as organizational citizenship performance.

The term Organizational Citizenship Behavior (OCB) was first proposed by Smith et al., (1983) in an effort to introduce non-task behaviors. Organ (1988) formally defined OCB as “behaviors of a discretionary nature that are not part of employees’ formal role requirements, but nevertheless promote the effective functioning of the organization” (p. 4). This conceptualization basically defines OCB as voluntary behavior that is not part of the job description. Organ (1988) identified five categories of OCBs (See Table 2). These five categories include behaviors that relate to helping coworkers, behaviors that contribute to the organizational environment, or behaviors that relate to being conscientious. Organ (1997) revised his original definition of OCB, as he noted problems with the term extra-role and acknowledged that some elements of OCB are appraised and likely to be rewarded. His revised definition includes discretionary behavior that contributes to organizational effectiveness.

Brief and Motowidlo (1986) introduced the concept of Prosocial Organizational Behavior in the late eighties in an attempt to evaluate the role of prosocial behaviors in organizations. They defined it as “behavior that is a) performed by a member of an organization, b) directed toward an individual, group, or organization with whom he or she interacts while carrying out his or her organizational role, and c) performed with the intention of promoting the welfare of the individual, group, or organization toward which it is directed” (p. 711). Brief and Motowidlo list 13 types of Prosocial Organizational Behaviors (See Table 2). Prosocial Organizational Behaviors include behaviors that are either functional or dysfunctional to the organization. The functional behaviors will be discussed in this section; the dysfunctional behaviors will be discussed under counterproductive behaviors. Rotundo (2000) notes that the functional behaviors

proposed by Borman and Motowidlo are similar to OCBs in that they include helping others, protecting the organization and its image, and individual effort.

George and Brief (1992) introduced the term Organizational Spontaneity to define behavior that is extra-role, performed voluntarily, and contributes to organizational effectiveness. They describe five forms of Organizational Spontaneity (See Table 2). Organizational Spontaneity also has direct parallels to OCB and Prosocial Organizational Behavior: helping coworkers and behaviors that benefit the organizational environment (Rotundo, 2000).

Borman and Motowidlo (1993) describe Contextual Performance as discretionary behaviors that apply across all jobs, are not necessarily role prescribed, and that contribute to the social and psychological environment of the organization. Borman and Motowidlo identify five types of contextual behaviors (See Table 2). A review of this list reveals that they can be grouped into three categories: helping others, helping the organizational environment and its image, and exerting effort and are comparable to Organ's, Brief and Motowidlo's, George and Brief's, and Borman and Motowidlo's conceptualizations.

Van Dyne, et al. (1995) defines Extra-role behaviors as "behavior which benefits the organization and/or is intended to benefit the organization, which is discretionary and which goes beyond existing role expectations" (Van Dyne, et al., p. 218). This definition requires that the behavior be non role-prescribed and not formally rewarded. Van Dyne et al. (1995) describe extra-role behavior on a continuum from Affiliative to Challenging and Promotive to Prohibitive producing a typology of Extra-role behaviors described by four groups (See Table 2). Two groups are similar to some of the behaviors presented in the other taxonomies: helping behaviors and behaviors that challenge a situation with the intent of improving it.

Organizational citizenship performance describes a unique set of behaviors that do not fall under task performance but contribute in a positive way to the organization. A number of researchers have acknowledged the importance of these types of behaviors in the workplace. However, the literature on organizational citizenship performance includes many different definitions and category schemes. In a review of this literature, Rotundo (2000) attempted to sort through these ideas and to devise a definition and category scheme that incorporate the essential elements of organizational citizenship performance, as conceptualized by these researchers. However, her goal was a definition and category scheme that are not situationally specific and that focus on the nature of the behavior itself without reference to whether it is in-role or extra-role and rewarded or not. She argued that restricting a definition to being in-role or extra-role or listed in the job description or not makes it difficult to compare organizational citizenship performance (or task performance) across organizations. With these criteria in mind, Rotundo concluded that organizational citizenship performance is nontask behavior that contributes to the goals of the organization by contributing to its social and psychological environment.

A review of the different category schemes included in Table 2 (See Table 2) indicates that a total of 32 groups of behaviors have been used to describe organizational citizenship performance. Rotundo (2000) argues that there is overlap among these categories, which can be grouped further into behaviors that are targeted towards coworkers (i.e., helping and informing others) or the organization (i.e., promoting the organization and its image) and behaviors that relate to an individual's determination, perseverance, or self-development. However, Rotundo argues that behaviors that involve persevering, determination, and self-development describe the manner in which tasks are performed and should not be included under organizational citizenship performance.

#### IV) Counterproductive Performance

Researchers have also written about nontask behaviors that have negative consequences for organizations and employees (e.g., Crino, 1994; Hollinger & Clark, 1982; Murphy, 1993; Raelin, 1994). These behaviors will be subsequently labeled counterproductive performance. As with the literature on organizational citizenship performance, there are numerous terms, definitions, and taxonomies that have been used to describe this group of behaviors (e.g., Crino, 1994; Hollinger & Clark, 1982; Raelin, 1994; Robinson & Bennett, 1995).

In a review of this literature, Robinson and Greenberg (1998) compared eight different conceptualizations of employee deviance against five defining characteristics: the perpetrator of the action, the intentionality of the action, the target of the action, the deviant action, and the consequences of the action. They reported the following trends: 1) The perpetrator was typically reported to be an insider; 2) Almost all definitions were restricted to include only acts that were intentional and voluntary; 3) The target toward whom the deviant acts were directed was often an individual who worked within the firm or the organization itself; 4) Deviant actions included violating organizational or group norms that brought harm to the target in one of a variety of different ways (e.g., direct, active, verbal, etc.); 5) The acts that were committed were harmful or had the potential to cause harm.

As a result of this analysis, Robinson and Bennett (1995) defined deviant behavior as “voluntary behavior that violates significant organizational norms and in so doing threatens the well-being of an organization, its members, or both” (p. 556). They stress that the behavior has to run counter to organizational norms in order for it to be considered deviant because they believe that deviance is specific to the standards specified by a social group, in this case the organization, and not moral standards. Rotundo (2000) notes that “threatening the well-being of

an organization” poses too rigid a restriction on this definition. That is, some behaviors satisfy all of the criteria in this definition and would be considered negative by members of an organization or the average person (e.g., continuously arguing with coworkers, difficult to get along with), but are not negative enough that they threaten the well-being of the organization. Hence, these types of behaviors would not be considered deviant by Robinson and Bennett’s standards, but would be considered deviant by the average person.

Robinson and Bennett (1995) used multidimensional scaling to develop a typology of deviant workplace behaviors. In their study, a group of participants rated pairs of behaviors for their similarity. Multidimensional scaling produced a typology of deviant workplace behavior that suggests that deviant acts can vary along two dimensions (i.e., interpersonal/ organizational, minor/ serious) and can be classified into four categories: *Property deviance*, *Production deviance*, *Political deviance*, and *Personal aggression* (See Table 2). *Property deviance*, borrowed from Hollinger and Clark (1982), represents serious acts committed at the level of the organization. It is defined as “those instances where employees acquire or damage the tangible property or assets of the work organization without authorization” (Hollinger & Clark, p. 333). The second category, *Production deviance*, also borrowed from Hollinger and Clark, represents less serious acts committed at the level of the organization. It is defined as “behaviors that violate the formally proscribed norms delineating the minimal quality and quantity of work to be accomplished” (Hollinger & Clark, 1982, p. 333). The third category, *Political deviance*, represents minor and interpersonal acts. Robinson and Bennett defined it as social interaction that puts other individuals at a personal or political disadvantage. The fourth category, *Personal aggression*, represents serious and interpersonal acts. They defined it as behavior that was



aggressive or hostile towards other individuals. This study represents one of few efforts to understand the structure underlying counterproductive behavior.

Gruys (1999) conducted a similar study in which individuals were asked to rate pairs of deviant behaviors for their co-occurrence. For example, individuals were asked the following question, “If an individual will engage in deviant behavior X, is that individual also willing to engage in deviant behavior Y?” Specifying the basis upon which individuals rate similarity makes for a clearer interpretation of the results. She found that behaviors fell along two dimensions. The first dimension ranged from personal (e.g., verbal comments) to impersonal (e.g., sabotage) deviant acts. The second dimension ranged from task (e.g., poor work) to non-task (e.g., misuse of resources) deviant acts.

Crino (1994) defines employee sabotage behavior as behavior intended to “damage disrupt, or subvert the organization’s operations for the personal purposes of the saboteur by creating unfavorable publicity, embarrassment, delays in production, damage to property, the destruction of working relationships, or the harming of employees or customers” (p. 312). In this definition the behavior is intentional and premeditated where the harmdoer is aware of the consequences. Crino believes that the type of sabotage varies by level of worker. Acts such as writing derogatory or threatening memos or damaging computer systems are more consistent with White-collar sabotage, where destroying machinery or tampering with a product is more likely with Blue-collar workers.

Raelin (1994) believes that organizational deviance is motivated by an employee’s experience with conflict or unacceptable work conditions combined with a need to stay with the organization (e.g., family responsibilities, benefits, no alternatives). Some employees choose to engage in deviant behavior as a way of dealing with this conflict. Raelin used Guttman scaling to

evaluate a model of professional deviant/adaptive behavior. Raelin proposed that these behaviors could be cumulatively scaled on a continuum according to four career elements: management, job, self, and career. He found support for a career, job, and self-scale (See Table 2). Career related adaptive behavior includes a premature external job search, where deviant behaviors include flaunting external offers. Job-related adaptive behavior includes bootlegging, where deviant behavior includes poor performance. Self-related adaptive behavior includes seeking outside interests where deviant behavior includes alienation from work.

Progress has been made in determining the structure underlying counterproductive performance. However, there are still unresolved issues (Robinson & Greenberg, 1995; Rotundo, 2000). Rotundo notes that a comprehensive definition of counterproductive or deviant job performance does not exist although some of the definitions come close. Furthermore, she notes that the lists of deviant acts that researchers have generated do not clearly delineate the boundaries within which deviant acts fall.

Researchers' interest in counterproductive behavior has been partially motivated by the desire to understand the underlying causes of this type of behavior. Research has shown that the determinants of deviant behavior include individual, social and interpersonal, and organizational factors (Robinson & Greenberg, 1998).

Concerning individual factors, a popular belief is that some people are more predisposed than others to engage in deviant acts. Rotundo (2000) notes that research on integrity tests as measures of one facet of personality provide some insight into individual difference variables as determinants of deviant acts. A large-scale meta-analysis of integrity test validities by Ones, Viswesvaran, and Schmidt's (1993) is one of the most significant contributions in this regard. Integrity tests measure honesty and moral character. This research showed that integrity tests

predict a variety of counterproductive behaviors (mean operational validity of .47), providing support for individual differences as determinants of counterproductive behavior.

Research shows that demographic characteristics are related to deviant behaviors (Frank, 1989; Hollinger & Clark, 1983). Although research has shown that young new-hires who work part-time in low-paying jobs are likely to engage in some forms of production and property deviance, these findings do not necessarily generalize across all job groups (Murphy, 1993). Research showed that perceptions of having been treated unfairly on the job were related to various types of deviant behavior such as retaliation (Skarlicki & Folger, 1997), stealing (Greenberg & Scott, 1996) and sabotaging work (Giacalone & Greenberg, 1997). There has been some evidence to suggest that underpayment inequity leads to theft (Greenberg, 1990) and aggression (Folger & Baron, 1996).

### ***Summary***

Progress has been made at understanding the domain of job performance. However, the field has further to go before the measurement of individual level job performance is fully understood. The different conceptualizations of job performance that were reviewed incorporate a variety of behaviors that can be grouped into three broad categories of behaviors: task, organizational citizenship, and counterproductive performance. Task performance was defined as behavior that relates to the production of a good or the provision of a service. The organizational citizenship domain was defined to include behavior that contributes to the goals of the organization in a positive way by contributing to its social and psychological environment. Behaviors include helping others, keeping others informed, promoting the image of the organization, volunteering, and making suggestions for improvement. The domain of counterproductive behavior was defined to include those behaviors that are voluntary, violate

organizational norms, and harm the well-being of the organization. Some examples of these behaviors include misuse of time and resources, inappropriate verbal actions, or poor quality work. Although these three categories describe the structure underling job performance, it is equally important to consider measurement issues related to individual level job performance.

# Measuring Job Performance

## *360-Degree Feedback*

Organizations' use of multi-source feedback can be traced back to the 1950s and 1960s, (Wood et al., 2000). Although multi-source feedback programs are extremely popular, they are not very well understood. Multi-source feedback incorporates performance appraisal from multiple raters who possess different perspective on employee job performance (Church & Bracken, 1997). A more commonly known name is 360-degree feedback. It has been estimated that about 40% of organizations use a form of 360-degree feedback (Bracken, Timmreck, & Church, 2000), thus highlighting the importance of research in this area.

An employee who receives 360-degree feedback is evaluated by raters who interact with or observe the target person (Wood et al., 2000). These people commonly include a supervisor, peers, self, subordinates, and clients. A feedback system such as 360-degree is based on the assumption that observations that are obtained from multiple sources yield more reliable and valid results, thus being more meaningful and useful for the individual (Church & Bracken, 1997). The multi-rater feedback obtained is intended to be an "instrument for personal development and performance improvement" (Jansen & Vloeberghs, 1999, p. 456), which, if used correctly, should result in positive behavior change. Although this assumption has not been entirely validated when applied to performance feedback, both practitioners and members of organizations believe strongly that multi-source feedback is effective.

Because of the strong practitioner belief that multi-source feedback is effective, it is crucial for research to demonstrate that it is effective. It is often the case that practices in organizations are perceived to be effective when they are not necessarily effective. There is a need for more investigation of the multi-source feedback area to clear muddled findings and to

provide more definitive answers.

A related issue surrounding multi-source feedback is whether the feedback should be used for developmental or evaluative purposes. Currently, many organizations use the multi-source feedback for developmental purposes. It may be difficult to base promotions and pay raises on peer, self, or subordinate ratings, as it could be argued that these ratings are influenced by many sources of bias that are unrelated to employee job performance. This becomes especially problematic if ratees do not accept the ratings or believe that they are not valid. That these situations may be challenged in the courts, explains in part why many organizations are hesitant to use 360-degree ratings to make such important decisions. Further research into this area could provide definitive answers about the purposes and outcomes of 360-degree feedback.

Questions about 360-degree feedback generally fall into four categories. These categories include the reliability and validity of 360-degree feedback ratings, employee reactions to 360-degree feedback, and the effect of 360-degree feedback on behavior change.

#### I) Reliability of 360-Degree Feedback

One problem widely studied in the area of multirater feedback is whether ratings provided by peers, subordinates, and supervisors are consistent with each other. Interrater correlations are thought to provide an estimate of the reliability of these ratings.

In an effort to investigate the reliability of 360-degree performance ratings, McEnery and Blanchard (1999) compared ratings of the same trait provided by different raters with ratings of different traits provided by the same raters. For example, they compared peers, subordinates, and self-ratings of leadership with peers' ratings of leadership, problem-solving, judgment, etc. Participants were 261 undergraduate business students in a required management skills course, who took part in a four-hour role-play simulation, where observers assessed behaviors

(assessors). The behaviors assessed included communication, leadership, managing the job, problem analysis, and teamwork. Self-ratings and anonymous peer ratings were completed after the simulation. The results indicated a lack of convergence for assessor-peer and assessor-self ratings. However, convergent validity and divergent validity (though to a lesser degree) did appear to be present in the peer-self ratings. Thus, McEnery and Blanchard's (1999) results indicate that different levels of raters do indeed emphasize different dimensions of performance, supporting the view that validity within groups is more valuable than determining rating validity between groups. They suggest that if convergence in ratings were a goal, then substantial training of raters would be necessary.

To see if ratings regardless of level captured unique rating variance, Mount et al. (1998) studied 2,350 managers from several industry groups to test five models with different factor structures of rating methods (raters) and traits (managerial skills) that were hypothesized to account for variance in performance ratings. Participants completed the Management Skills Profile (MSP), which contained 116 items representing 18 dimensions of managerial behavior. The MSP had three managerial skill categories: administrative, human relations, and technical competence. Seven ratings (self-ratings, two subordinates, two bosses, two peers) were available for each person rated. The first model hypothesized that covariation in performance ratings was associated only with traits of the ratee (managerial skills) and not with the rating method (rater). The second model hypothesized that the covariation in ratings was thought to be associated with individual raters and not with the traits of the ratees. A third model included some trait and some rater effects. The fourth model hypothesized both trait and rater effects. Using confirmatory factor analysis, Mount et al. (1998) found that the best fit for the model that hypothesized 7 rater and 3 trait effects. These results indicate that method variance in

performance ratings is associated more strongly with individual raters than with the level of the ratings. Because the ratings from each rater capture unique rating variance, their results support the notion that multiple raters are important in the appraisal process. Their findings also suggest that aggregating ratings from similar-level raters is inappropriate, because these ratings are no more similar to each other than ratings from different-level raters. Aggregating performance ratings appears to lower the reliability of the ratings, though it does increase the estimated reliability of the ratings, according to Mount et al. (1998).

Personality is one area that has been explored as a moderator of self-other agreement. Smith and Helland (2002) conducted a study to examine whether personality traits such as self-acceptance, dominance (self-confidence and feelings of security), and flexibility predict whether self and other raters agree. They also considered correlates of self-other agreement, such as the effect of managers' overestimation of ratings on their leadership and judgment. Analytic ability was also observed in relation to the above variables. Studying 223 executive MBA students, they asked each participant to provide three to five subordinate ratings and to rate themselves. The results showed that personality traits such as dominance and flexibility were significant predictors of agreement; that is, individuals who have dominant and inflexible personality traits were most likely to overestimate themselves relative to their subordinate ratings on leadership behavior measures. Analytic ability only partially positively predicted self-other agreement, and self-acceptance did not predict agreement at all.

Agreement has also been studied from the subordinate perspective. Johnson and Ferstl (1999) examined the effect that feedback from subordinates would have on managers' self-ratings. They also differentiated between managers who are considered to be over-raters (those who rate themselves higher than others do), and under-raters (those who rate themselves lower



than others do). Participants in the study included 902 partners in an accounting firm, 732 senior managers, and 415 managers. Over approximately two years, managers completed self-ratings and had up to ten subordinates complete performance ratings on the manager. Results indicated that managers who were considered to be under-raters or in-agreement raters (ratings were consistent with subordinate ratings) did not increase their self-ratings following upward feedback from subordinates. Managers who rated themselves lower than subordinates showed a decline in ratings of their own performance over time. Subordinate ratings of over-raters increased following feedback, where subordinate ratings of under-raters and in-agreement raters remained constant following feedback.

Some patterns emerge in the findings presented so far. Performance tends to increase when self and other (specifically, subordinate) ratings are high and in-agreement (Atwater et al., 1998; Halverson et al., 2002). However, Johnson and Ferstl (1999) found some contradictory evidence, suggesting that subordinate ratings of overraters increased following feedback. Still, these findings generally replicate work done previously in this area (Atwater, Roush, & Fischthal, 1995; Atwater & Yammarino, 1992; Van Velsor, Taylor, & Leslie, 1993).

Unreliability in ratings can often be attributed to rating errors, or errors in judgment that may arise when one individual is observed and rated by another (Latham & Wexley, 1994). One such error is the leniency error, when a rater gives a too-harsh or too-indulgent rating. Antonioni and Park (2001) investigated the influence of rater affect on the leniency of ratings. In particular, they hypothesized that in downward, upward, and peer feedback, as rater affect towards ratees increased, so would raters' leniency towards ratees' work behaviors. Antonioni and Park also hypothesized that the influence of rater affect on rating leniency would be stronger in both peer and upward ratings rather than in downward ratings, and that the influence of rater

affect on rating leniency would increase as the rater observed the ratee more often. A data set was used consisting of 163 downward ratings, 103 upward ratings, and 1027 peer ratings of people in all positions at a midsize insurance company. Results indicated that a rater's interpersonal affect does influence the three sources of 360-degree feedback, the influence being stronger in upward and peer ratings than in downward ratings. This underlines the theme that raters differ by level in how they give feedback. Also, the more frequently raters observed ratees, the more the influence of rater affect of ratings increased. These findings provide support for current efforts to train raters to avoid certain biases such as the halo or leniency effects.

Furnham and Stringfield (1998) studied congruence in job performance ratings, also focusing on the leniency effect. Specifically, they hypothesized that participants' self-ratings would be higher than those of peers, managers, or consultants. Self-other correlations were hypothesized to be lower than other-other (manager-peer, manager-consultant, peer-consultant) correlations. To test these hypotheses, Furnham and Stringfield studied seven teams of one manager, six peers, and an outside consultant over the period of one year. Managers and consultants rated each member of the team at the end of the year. Each team member rated himself or herself along with his or her peers. Evidence of a leniency effect was found, with participants rating themselves higher than others rated them. Self-other correlations were found to be lower than other-other correlations, confirming the hypothesis. The findings in this study provide support for previous studies that have examined self-other and other-other ratings (Harris & Schaubroeck, 1988; Furnham & Stringfield, 1994). However, the other-other correlations in this study were much higher than in comparative studies, while self-other correlations were very lower.

Farh, Werbel, and Bedeian (1988) examined the correlations between self-ratings and supervisory ratings by comparing 88 full-time faculty members' self-ratings of performance with the chairperson ratings of their performance. The performance ratings were based on seven dimensions: journal publications, instructional method, instructional support, meeting presentations, university and college service, department service, and professional service. Self-ratings were found to be highly congruent with supervisory ratings, and self-ratings were no more lenient than supervisor ratings. These results present a different picture than the findings of Furnham and Stringfield (1998). These differences may again be due to the lack of independence in ratings in the Farh et al. study, where supervisor ratings were received only after the chairpersons had reviewed the self-ratings. Also, the process in this study used a common pool of information about performance for the self and chairperson ratings. Rater training such as frame-of-reference training, which is not usually done in 360-degree programs, could possibly reduce leniency in ratings. This would greatly enhance accuracy (Waldman et al. 1998; Wood, Pillinger, & Kohn, 2000).

Viswesvaran, Ones, and Schmidt (1996) compared the reliability of performance ratings of different dimensions in a meta-analysis. They included studies that reported reliability coefficients for overall job performance, job performance or productivity, quality, leadership, communication competence, administrative competence, effort, interpersonal competence, job knowledge, and compliance with or acceptance of authority. For supervisor ratings, Viswesvaran, Ones, and Schmidt found a mean interrater reliability of .52 for overall job performance. The interrater reliability of overall job performance ratings was similar to the mean interrater reliability across job performance dimensions for supervisors. Overall job performance reliability for supervisors was .86. Peer interrater reliability of overall job

performance was somewhat lower than that of supervisors at .42., and identical to mean interrater reliability for peers. Also, for peers, the ratings of overall job performance were similar to those of supervisors, at .85. For both peer and supervisory ratings and for all dimensions and overall ratings, interrater reliability estimates were substantially lower than intrarater reliability estimates. Twenty to 30 percent of the variance in job performance dimension ratings of the average rater was found to be specific to the rater.

## II) Validity of 360-Degree Feedback

Research that examines the validity of 360-degree feedback seeks to determine whether or not performance ratings from peers, subordinates, or the self predict organizational success, ratings of overall job performance, promotion rate, pay-raises, or other outcomes. The following paragraphs discuss studies that examine different types of validities.

Halverson, Tonidandel, Barlow, and Dipboye (2002) studied whether agreement between raters predicts promotion rates. Agreement in this study was compared using self-ratings on a leadership measure and other-ratings (subordinates, supervisors, and peers) on the same measure. They hypothesized that in-agreement raters (those raters whose self-ratings are consistent with others' ratings) would be the highest performing employees, followed by under-raters and over-raters, respectively, for all organizational others. Under-raters and over-raters are those who either underestimate or overestimate their ratings relative to others' ratings. Additionally, the relationship between self-subordinate agreement was hypothesized to be important for predicting future performance. Using a 1,500-person sample from the US AirForce (lieutenant, captain, major, or lieutenant colonel), Halverson et al. administered a 360-degree rating instrument that focused on leadership skills. Analyses revealed that self-subordinate agreement was more important than self-peer or self-supervisor agreement in predicting promotion rate. Halverson et

al. also found that those participants high in agreement had the highest promotion rate. In summary, self-awareness (the consistency between self and other ratings) contributes highly to the prediction of the outcome measure, promotion. The agreement of self and subordinate ratings may be important when self and subordinate ratings are based on leadership behavior.

Sala and Dwight (2002) examined whether multi-rater feedback predicted top-level corporate executive performance. Participants were 276 senior executives from a global technology company. Participants were assessed over a four-year period on competency behaviors (organizational leadership, managerial leadership, business leadership, and intrapersonal leadership). The executives' performance was measured against the extent to which managers met their annual performance targets. Sala and Dwight (2002) report that direct report and manager ratings predicted job performance for corporate executives

Performance is often measured in more than one way; that is, constructs are often measured using both objective and subjective indicators. Bommer, Johnson, Rich, Podsakoff, and Mackenzie (1995) performed a meta-analysis to assess the relationship between subjective and objective performance measures. Subjective measures are considered to be supervisor ratings of employee performance, while objective measures are considered to be "direct measures of countable behaviors or outcomes" (Bommer et al., 1995, p. 588). Performance scores as assessed by objective measures did not correlate highly with performance scores assessed by subjective measures, indicating that they may not be used as substitutes for each other or be interchangeable when considering performance. However, when the analysis was restricted to samples that draw on the same construct (what) and compared objective versus subjective indicators (how), it appeared the measures were more substitutable when analyses were limited to samples that considered the same construct (the "what" was held constant) and

then compared objective and subjective indicators (the “how” differed). Still, Bommer et al. (1995) do not support the interchangeable use of objective and subjective measures.

### III) Reactions to 360-Degree Feedback

The measurement of user reactions to a 360-degree feedback system is vital to the success and viability of the feedback process (Edwards, Ewen, & Vendantam, 2001). User reactions and attitudes toward feedback have been studied using various indicators of reactions such as acceptability, satisfaction, and perceptions of usefulness.

Fedor and Bettenhausen (1989) investigated whether the acceptance of a peer evaluation system was affected by the following variables: the purpose of the peer ratings, whether peer evaluations were used for feedback, and the level of peer ratings received. The favorability of the participants’ preconceptions about the system’s usefulness and legitimacy were also evaluated. In their study, 208 undergraduates in an introductory management class were given peer feedback in feedback circles. In the first condition, peer evaluations were used for feedback only. In the second condition, the feedback ratings accounted for 50% of the students’ grade for class participation. Feedback was collected and given to participants twice during a university term. Fedor and Bettenhausen found a significant effect for the purpose of the peer ratings, and interestingly, that the acceptance of peer evaluation was higher when peer evaluations were used for grading purposes. Also, those who were rated higher by peers were less likely to accept the peer evaluation system as beneficial than those who were rated less favorably. Finally, results indicated that participants’ preconceptions about the evaluation system were significant predictors of participants’ final acceptance of the system. This may explain why employees do not always accept feedback as useful or relevant: If they have unfavorable views on the feedback system before it is in use, their acceptance and use of the system will not necessarily provide

accurate or valuable information for other people, be they peers or subordinates. Still, whether an employee finds a certain feedback system acceptable does not imply that the evaluation system will be used fairly or accurately. Other factors may play into the situation.

In another study investigating acceptance of feedback, Nease, Mudgett, and Quinones (1999) used self-efficacy as a predictor of feedback acceptance. Specifically, they hypothesized that in conditions of negative feedback, individuals with high self-efficacy would display decreased feedback acceptance, while individuals with low self-efficacy would not display a change in feedback acceptance. Their second hypothesis invoked the use of feedback acceptance as a moderator; that is, individuals with high acceptance ratings who received positive feedback would be more likely to subsequently increase in self-efficacy than those with lower ratings, and when receiving negative feedback would be more likely to subsequently decrease in self-efficacy compared to those with lower ratings. To test their hypotheses, the researchers asked 80 undergraduate psychology students to complete a computer simulation. Participants received feedback in two conditions: They performed either 20% below the goal in the negative feedback condition or they performed 20% above the goal in the positive feedback condition. Measures of self-efficacy were completed before and after the trials for the computer simulation. The first hypothesis was confirmed: High self-efficacy participants did display a significant decrease in acceptance ratings of the feedback, in conditions of negative feedback. The second hypothesis was only partially supported: Acceptance of feedback moderated the relationship between feedback sign and the predication of self-efficacy during later trials, though not at first.

Recently, Hedge and Teachout (2000) collected data on rater attitudes about acceptability of performance appraisal forms in an attempt to extend the development of an acceptability criterion. Their participants included 1608 job incumbents in their first four years of military

service, in seven skilled trade and clerical jobs in the air force. Five hundred and thirty-four of their supervisors also participated. Data were collected over three years. Analyses revealed that rater motivation, rater trust, and situational constraints influence job incumbent and supervisor acceptability of the appraisal process. Situational constraints such as internal and external work impediments (for example, problems with equipment or policies) were found to interfere with the ability to judge proficiency fairly, accurately, and confidently. Differences in levels of acceptability were also found, with global ratings being significantly less acceptable to all raters. Another interesting finding was that the supervisor perceptions of the appraisal system were more favorable than incumbent perceptions.

Related to whether employees accept 360-degree feedback is whether employees are satisfied with the performance appraisal and the appraisers. Boswell and Boudreau (2000) addressed this question by investigating whether employee perceptions that the performance appraisal were used for development was positively associated with employee satisfaction with both the performance appraisal and appraiser. Participants included 128 employees at a production equipment facility who were given performance appraisals by supervisors once a year. Employee attitudes two months after the appraisal were assessed. Perceived performance appraisal use for development was positively related to both performance appraisal and appraiser satisfaction. However, the performance rating did not moderate the relationship between perceived use for evaluations and reactions to either the appraisal or appraiser.

Another employee reaction relationship that is important in the appraisal process is whether employee participation in the appraisal process is related to employee reactions to the performance appraisal. Cawley, Keeping, and Levy (1998) performed a meta-analysis that compared subordinate reactions (satisfaction, motivation to improve, utility, and fairness) to



participation in the performance appraisal process (value expressive, instrumental, and time talked). Value-expressive participation is participating in order to have one's voice or opinions about the appraisal heard, whereas instrumental participation is participating in order to influence the end result of the appraisal process. The results from the meta-analytic review indicate that there is a strong relationship between performance appraisal participation and subordinates' affective reactions. The value expressive participation and instrumental participation (to a lesser degree) were strongly related to overall reactions, with value expressive participation more strongly related to positive reactions than instrumental participation. Thus, there seems to be a strong, stable relationship between participation in the appraisal process and appraisal satisfaction.

If employees perceive feedback to be accurate, it seems to be likely that they will also perceive feedback to be useful. Brett and Atwater (2001) examined this issue, as well as positive and negative reactions to feedback. Specifically, they investigated whether perceptions of feedback accuracy were positively related to positive reactions, and negatively related to negative reactions. To test these hypotheses, Brett and Atwater studied 125 MBA students over a semester. Participants received feedback from their previous employers, attended a training session for interpreting feedback, and later met with a facilitator to discuss developmental plans. Reactions and attitudes to feedback were also assessed several times. The results revealed some interesting findings for positive and negative reactions. The feedback from bosses and direct reports were rated more accurate if the feedback was more positive (although this did not hold true for peer ratings). When participants rated themselves higher than the boss or direct reports, they perceived the boss and direct reports to be less accurate (this did not hold true for self or peer ratings). Negative feedback was not viewed as accurate or useful, and rather than resulting

in awareness, resulted in several negative reactions, such as anger and discouragement.

However, high ratings were not related to positive reactions, but rather to the absence of negative reactions.

Bettenhausen and Fedor (1997) compared peer and upward appraisals, looking at variables such as appraisal outcomes, coworker and supervisor relations, and experience with peer and upward appraisals. Participants included 195 people from graduate classes or the business community. The research design was a 2-appraisal type (upward vs. peer) by 2-appraisal purpose (administrative vs. developmental) within-subjects factorial design, where participants completed measures of the above variables. Results indicated that employees viewed appraisals used for developmental purposes more positively than they did appraisals used for administrative purposes. Participants believed both positive and negative outcomes would result when upward appraisals were used as input for administrative decisions. However, they believed negative outcomes would result and positive outcomes were not likely if peer appraisals were used as input into administrative decisions. The quality of coworker relationships affected participants' reactions to appraisals, but the quality of relations with their supervisors did not.

Attitudes toward 360-degree feedback systems may be influenced by individual and contextual variables. Funderburg and Levy (1997) hypothesized that individual factors (i.e., self-esteem, tendency to seek feedback, and locus of control) and contextual factors (i.e., organizational citizenship behavior, feedback-seeking environment) would be positively related to attitudes toward 360-degree feedback systems. Other contextual factors were also hypothesized to be related to 360-degree feedback attitudes: a) Perceived costs associated with feedback seeking were thought to be negatively related to attitudes toward 360-degree feedback systems, and b) supervisory style were thought be related to attitudes toward 360-degree

feedback systems such that those employees with more autocratic supervisors would be more favorable toward 360-degree feedback systems. Participants in this study were 75 employees from manufacturing and telecommunications organizations who had been involved in a multi-rater performance appraisal system within one to two years prior to this study. They were asked to complete the measures for individual and contextual factors, and three coworkers of each employee completed an organizational citizenship behavior measure. Correlational analyses revealed that as a whole, predictor variables accounted for nearly 50% of variance in multi-rater appraisal system attitudes. Self-esteem, locus of control, and a feedback seeking environment were positively related to attitudes toward 360-degree feedback systems, but tendency to seek feedback was not significant. Furthermore, organizational citizenship behavior was not significantly related to attitudes. However, perceived costs were negatively related, and e) supervisory style was significantly related to attitudes. In summary, contextual factors were more important (explained more variance) than personality factors in determining attitudes toward 360-degree feedback.

#### IV) Effectiveness of 360-Degree Feedback

The question that should be most crucial when deciding to use 360-degree feedback is, does it work? That is, does behavior change in a positive direction when 360-degree feedback is used? The following studies provide some evidence to believe that multisource feedback works in certain situations and at certain times.

Atwater, Ostroff, Yammarino, and Fleenor (1998) explored the relationship between self, peer and subordinate ratings, and perceptions of managerial effectiveness. Participants in the study were 1460 managers who participated in a leadership development program. Managers, peers, and subordinates completed the 106-item multi-rater feedback instrument, are the

supervisor completed the 16-item managerial effectiveness measure. Hierarchical regression analyses revealed that ratings of managerial effectiveness tended to be higher when both self and other ratings were high than when they were low. Ratings of managerial effectiveness were lower when self-ratings were greater than other ratings. Thus, Atwater et al. (1998) conclude that it is important to consider self-ratings and other ratings when measuring outcomes such as managerial effectiveness. This research indicates that self and other ratings do predict managerial effectiveness but when both ratings are high.

Dominick, Reilly, and McGourty (1997) investigated whether giving and receiving feedback influences team behavior. Seventy-five graduate and undergraduate students in management were organized into teams of four or five, and then randomly assigned to feedback, exposure, or control conditions. They then performed two counterbalanced group decision-making tasks. In the feedback condition, participants completed self and peer ratings after the first task and received feedback prior to the second task. In the exposure condition, participants completed self and peer ratings after the first task and were only given feedback after the second task, whereas in the control condition, participants did not complete self or peer ratings and received no feedback. Results indicated that participants who either gave and received feedback, or gave but did not receive peer feedback, displayed effective team behavior more often than participants who did not give or receive peer feedback. Finally, participants who gave and received feedback did not differ in effective team behavior from participants who gave feedback only. This indicates that both giving and receiving of peer feedback contributes to more effective team behavior, with the giving of feedback being just as effective at behavior change, if not more so, than receiving feedback from peers. It could be argued that exposure to the feedback instrument focused team members' attention on the important behaviors needed for

effective performance and conveyed certain valued and important behaviors to the team members.

Druskat and Wolff (1999) examined the effects of developmental peer appraisals on self-managing workgroups. Eighty self-managing project groups were created from 511 undergraduate and graduate students. Questionnaires measuring perceptions were administered four times over the course of one university term. Druskat and Wolff found that structured, face-to-face developmental peer appraisal had a positive impact on group member perceptions of open communication, group task focus, group viability, and member relationships. These findings occurred for both graduate and undergraduate students. The positive perceptions on communication, focus, and relationships may foster increased task motivation and ability to complete the task. Another interesting finding was that lasting development resulting from the appraisal depended on the timing of the peer appraisal relative to the group's deadline for completing a project.

Accountability is thought to be an important part of a feedback process (London, Smither, & Adsit, 1997). Raters should be accountable to ratees to provide "accurate, meaningful information" and ratees should be held accountable to use the information (London, Smither, & Adsit, 1997). Therefore, Smither, London, Reilly, Flautt, Vargas, and Kucine (2002) asked whether the action of discussing the multisource feedback with the raters would enhance performance. They hypothesized that ratees who shared their multisource feedback with raters and asked for suggestions would improve more than ratees, who did not share their feedback. Participants were 5,335 ratees in a global corporation who received multisource feedback three times in a one-year period. Hierarchical regression analyses revealed that sharing

feedback and asking for suggestions accounted for a very small proportion of variance in improvement over time, albeit a significant portion.

Walker and Smither (1999) also used the concept of accountability when they investigated managers' performance over five years. They hypothesized that managers' performance would improve over the five years, and that managers who initially received the poorest ratings would improve more than managers who initially received more favorable ratings. They also hypothesized that managers who met with direct reports to discuss the previous year's feedback results would improve more than managers who did not meet with direct reports. Participants included 252 managers who received feedback from subordinates at least three times over five years. Managers' performance was found to improve over time. Walker and Smither's hypotheses were confirmed in that managers who initially received the poorest ratings did improve more than managers who initially received more favorable ratings. Also, managers who held feedback sessions to discuss feedback improved more than other managers, and managers improved more in years when they held feedback sessions. Thus, discussing feedback after receiving it may be a valuable tool for behavior change.

Goal-setting in addition to feedback is thought to provide longer-lasting behavior results than just feedback alone (Latham & Locke, 1991). Brutus, London, and Martineau (1999) hypothesized a connection between 360-degree feedback and managers' planning for career development. Specifically, they hypothesized that supervisor ratings would yield the greatest influence on managers' goal selection followed by ratings from peers, direct reports, and self, respectively. Additionally, managers were thought to select developmental goals that were conceptually related to those specific performance dimensions in which they had received low ratings. The participants in the study included 2,163 managers who participated in a leadership

developmental program. Self, supervisor, peer, and subordinate evaluations were collected for every participant, and the results of this 360-degree evaluation was given to the managers at the beginning of the program, along with an individual feedback session. Managers completed a standardized goal-setting instrument after receiving the feedback. Contrary to the hypothesis, supervisory ratings were found to be the least important source of performance information; the most important rating sources were subordinates, peers, then self, respectively. Managers were found to focus on those dimensions in which they were rated lower when they set their developmental goals. Also, performance ratings and the goal selection relationship was stronger for lower level managers than higher level managers.

Smither et al. (1995) studied the effects of an upward feedback (feedback from subordinates) program on managers' performance. They hypothesized that managers would improve their performance or alter their self-evaluations after the implementation of an upward feedback program, and that these changes would be related to receiving written feedback, completing self-ratings, and the managers' initial level of performance. This study consisted of two waves, with 1,522 managers in the first wave, and 1,605 in the second. A survey consisting of 33 behavioral statements reflecting boss/subordinate relationships was administered to managers for the first wave, and six months later for the second wave. Results indicated that managers did improve their performance six months after an upward feedback program was implemented; however, performance improvement was observed regardless of whether managers received an upward feedback report.

Atwater, Roush, and Fischthal (1995) considered the influence of upward feedback on self- and subordinate leadership ratings. They hypothesized that the overall effect of feedback would be improvements in behavior as measured by follower ratings of leadership and

improvements in self-ratings. They also hypothesized that supervised practice would improve self-evaluations following feedback to a greater extent than feedback alone. Furthermore, they hypothesized that leaders receiving negative feedback (follower ratings substantially below self-ratings) would change their subsequent behavior while those receiving follower ratings in agreement with or higher than self-ratings would not change their behavior. Participants included 978 juniors and 1232 freshmen in the U.S. Naval Academy. Followers (freshmen) were required to provide feedback to the leaders (juniors) about their leadership. The followers completed a survey eight weeks and 14 weeks into the semester. At 16 weeks, leaders rated their own behavior toward followers. Some juniors were also required to have supervised practice to see that positive counseling sessions were taking place. Finally, at 32 weeks all followers rated the leaders. Atwaters, Roush, and Fischthal found that feedback about one's behavior from others tended to promote generally positive changes in leaders' behaviors and self-ratings. Supervised practice was not found to augment the effects of feedback alone. Leaders who received feedback from followers that their leadership was better than self-ratings significantly raised their self-ratings following feedback. Finally, leaders who overestimated behavior greatly improved their leadership. These results suggest that feedback received from others about one's behavior seems to encourage mostly positive changes in a leader's behaviors and self-ratings.

Another longitudinal study was done by Reilly, Smither, and Vasilopoulos (1996). They examined four administrations of upward feedback over two and a half years, to determine whether initial gains in performance were sustained or increased over time and whether performance improvement was related to the frequency of upward feedback. Using 92 managers, they collected data as part of an upward feedback program, which used four ratings at different times: July of 1992, January of 1993, January of 1994, and January of 1995. Measures



included a survey consisting of 33 behavioral statements reflecting boss/subordinate relationships and personalized feedback following the ratings. Results revealed that managers' whose initial level of performance was low had improved, and that the amount of improvement was not related to the number of times (or when) the managers had received upward feedback. Furthermore, they found that most of the performance improvement occurred between the first and second administrations of feedback and initial performance improvements were sustained (and slightly enhanced) over time. This suggests that an upward feedback program conducted over a period of time may result in sustained behavior change. Exposure to the valued behaviors may be more useful for sustained change than receiving feedback.

Atwater, Waldman, Atwater, and Cartier (2000) used a field experiment to determine supervisors' cynicism, leadership, and commitment to subordinates. Participants included 96 supervisors in a state police agency. In one condition, a feedback group received feedback at two time periods separated by ten months. In another condition, a survey-only group had supervisors and subordinates complete surveys at two times, with supervisors receiving feedback at Time 2 only. The survey consisted of 43 leadership items, eight organizational cynicism items, and three items for supervisor commitment to subordinates. The survey-only condition was intended as a comparison group, as the act of surveying was not thought to have an impact on the variables under study. Atwater et al. found that the feedback group did not show improvement in leadership from Time 1 to Time 2 compared to the survey-only group. Those who received feedback lowered their self-ratings but those who were surveyed only did not lower their self-ratings. Those who were more cynical about organizational change were less likely to have high ratings at Time 2. Those individuals who found feedback more valuable and were more likely to set improvement goals had higher scores at Time 2. Finally, lower ratings

from subordinates resulted in lower subsequent commitment to subordinates while higher ratings from subordinates resulted in higher commitment to subordinates. Thus, self-awareness and commitment to subordinates may be outcomes other than performance that are affected by upward feedback.

### *Summary*

Fortunately, a body of literature is emerging that examines important issues related to 360-degree feedback. However, the research results are not always consistent thus leaving room for future research to resolve the inconsistencies, to address study limitations, and to address further questions. The review of 360-degree feedback focused on four broad areas including reliability, validity, user reactions, and behavior change.

The research thus far on the reliability of ratings provided by peers, the self, subordinates, and supervisors indicates support for both the consistency of ratings and disagreement between raters of different perspectives. Although Farhl, Werbel, and Bedeian (1988) found that self-ratings are congruent with supervisory ratings of job performance, Furnham and Stringfield (1998) found that the agreement between self-ratings and ratings provided by raters from other perspectives is lower than the agreement between two ratings that do not include self-ratings. Furthermore, McEnery and Blanchard (1999) found that ratings provided by peers are consistent with ratings provided by the self and inconsistent with ratings provided by raters from other perspectives. The review also revealed methodological flaws in some of the studies. Therefore, future research needs to address these limitations.

A much more limited research base has examined whether 360-degree performance ratings predict organizational level outcomes such as promotion rates. The research that does exist shows that promotion rates are higher when both self and other ratings are high than when

they are low (Halverson, Tonicandel, Barlow, & Dipboye, 2002). Furthermore, research shows that subordinate and manager ratings predicted job performance for corporate executives (Sala & Dwight, 2002). There is room for future research to contribute to this literature base by examining whether 360-degree ratings predict additional organizational outcomes such as pay rates, merit increases, or voluntary turnover.

Although organizations seem to value multi-rater feedback, do the recipients of the feedback respond favorably to it? Research indicates that recipients' perceptions that multi-rater feedback is used for developmental purposes leads to increased satisfaction with the appraisal and appraiser (Boswell & Boudreau, 2000). However, when the feedback is negative, recipients perceive the feedback as low in accuracy or usefulness (Brett & Atwater, 2001). Furthermore, when the motivation for multi-rater feedback is framed as an opportunity for voice in the performance appraisal process, the reactions are more favorable than when the motivation is framed as an opportunity to influence the outcomes (Cawley, Keeping, & Levy, 1998). Future research needs to examine additional factors that moderate the acceptance of 360-degree feedback (e.g., timing). Other conditions under which these programs are implemented are likely to determine whether or not the recipients accept the ratings.

The ultimate goal of providing ratings from multiple perspectives is that as a whole these ratings provide useful information for the recipients who then modify their behavior in a positive way and improve job performance. Once again, research on this front produces mixed findings. Research indicates that feedback from subordinates improves leader behavior (Atwater, Roush & Fischtal, 1995), influences goal setting behavior (Brutus, London, & Martineau, 1999), and improves job performance over time (Walker & Smither, 1999). However, other research shows that various factors moderate this relationship. For example, managerial effectiveness increased

only when there was agreement between self and other ratings and not when self-ratings were lower than ratings provided by raters from other perspectives (Atwater, Ostroff, Yammarino, & Fleenor, 1998). Research also showed that either giving or receiving feedback produced an increase in effective team behavior compared to a control condition where feedback was neither given nor received (Dominick, Reilly, & McGourty, 1997).

For a period of time, researchers and practitioners focused mainly on supervisory ratings of task performance or objective indicators of results based measures of performance (i.e., quarterly sales or number of widget produced per shift) to determine or assess individual level job performance. The review thus far has suggested that this approach to measuring job performance is outdated at best. The review indicates that perhaps individual level job performance should incorporate behavioral based measures of task, organizational citizenship, and counterproductive performance that include the peer, self, subordinate, and supervisory perspectives. However, research to date has not answered all questions related to these issues and additional research is needed.

### ***Stability of Job Performance***

Part of process of understanding job performance is determining whether or not an individual's job performance changes over time or whether it remains stable from one month or year to the next month or year. Intuition suggests that job performance should improve with training, work experience, and mentoring/coaching. However, it is not clear whether or not individual level job performance changes or fluctuates even after controlling for these factors.

The stability of job performance has implications for the validation of selection tests. Companies decide to use employment interviews or ability/personality tests as part of their selection systems because they believe that these instruments provide information that will help them hire candidates who will perform well on the job. In order to be able to make assertions about the effectiveness of selection tests, these tests are validated against individual level job performance. Therefore, decisions about whether or not selection tests are valid rely on the assumption that individual level job performance is stable over time, otherwise, the validity of selection tests fluctuates if this assumption is not true. Thus, hiring and selection decisions are affected by the performance stability assumption, as these decisions are based on the theory that it is possible to predict long-term performance from indicators like test scores or early job performance (Hanges, Schneider, & Niles, 1990). It has also been suggested that the assumed relationship between ability and performance may not exist if performance does indeed change from one time to another (Schmitt & Chan, 1998). A finding that job performance is not stable makes the whole validation process difficult to justify.

### ***Approaches***

One approach to conceptualizing job performance is to view it as constantly changing rather than being stable (Ghiselli, 1956). This approach views job performance as somewhat

unpredictable and dependent on many factors such as the job, individual mood, or other factors. This approach labels job performance as “dynamic criteria”. Change in performance can be attributed to many things, one of which is the acquisition of new skills on the job. This means that performance may alter with learning and development on a job, as observed by Ghiselli (1956). On the other hand, if criteria are static, then it can be assumed that performance is stable over time. From the dynamic criteria viewpoint, certain predictors of an employee’s behavior, such as ability, may not be consistent throughout the employee’s tenure (Borman, 1991).

It is one thing to conceptualize job performance as dynamic, it is another to measure or prove that it is dynamic. There are different approaches to showing that job performance is dynamic. Dynamic criteria can be shown by changes in a group’s average performance over time. For example, some researchers computed the average job performance for a group of employees over a period of years and examined the data for any fluctuations in the group’s mean performance. This definition is considered to be conceptually the weakest, as research does not distinguish between performance during a training period and continuous improvement over a period of time after the training. Another weakness of this approach is that average curves computed based on mean group performance do not reveal changes in individuals’ relative positions.

Another approach to assessing whether or not job performance is dynamic or stable is by assessing individual level changes in the rank-ordering of their job performance scores over time. This approach involves a comparison of correlation coefficients computed between columns of job performance scores collected over time. For example, imagine a sample of 100 employees whose job performance is rated every week for one year. Each column of data represents job performance ratings for one week. In this approach, the correlation between Week

one and Week two is compared to the correlation between Week one and Week three. All possible correlations are computed and compared. This process enables one to assess change in rank ordering over time. A finding that the correlations coefficients are relatively similar provides support for the stability of job performance, where a finding that the correlation coefficients fluctuate over time provides support for the notion that job performance is dynamic. Although researchers argue that this definition is the only one that directly addresses the assumption of stability of performance, this approach and the research conducted using this approach is not without its limitations (Barrett et al.). In several instances change over time was found (hence support for dynamic criteria). However, other factors such as restriction of range and temporal unreliability may also explain the results. Hence, researchers need to improve criterion measure reliability.

In an extension of the rank-ordering approach, a group of researchers argue that dynamic criteria can be shown by a simplex pattern in the correlations (Austin, Humphreys, & Hulin, 1989). This pattern occurs when the highest correlation is observed between adjacent time periods (Time + 1) with systematic reductions in the magnitude of correlations as the interval between occasions increases (Time + 2, Time + 3, . . . , Time +  $k$ ) (Hofmann, Jacobs, & Gerras, 1992; Ackerman, 1989). The conclusion drawn from the observation of a simplex-like pattern is that individuals are changing their rank order over time, and thus that job performance is dynamic. Early studies conducted by Rothe (1978) and Ghiselli and Haire (1960) found that individuals do in fact change their rank order over time. Since these initial studies, many studies examining change in performance over time have discovered the simplex pattern.

Henry and Hulin (1987) have been strong proponents of the dynamic criteria concept, using the simplex pattern as evidence. Henry and Hulin's results indicated that the correlations

between ability and performance decrease as a function of the time between performance assessments. A meta-analysis conducted by Henry and Hulin (1990) found that predictive validity decreased over time. With few exceptions, Henry and Hulin found that these validities decrease monotonically. The following studies also provide support for the simplex pattern.

Rambo, Chomiak, and Price (1983) studied sewing machine operators over a three-and-a-half year period to investigate the long-term reliability of work performance. The sample consisted of 46 workers who were either sewing machine operators or who performed miscellaneous tasks such as folding and packaging the garments. Both jobs were highly specialized and routine with little variation of tasks. Rambo et al. (1983) obtained objective data on output rates from the workers' weekly production earnings report (the average hourly output for each week multiplied by the unit pay for each worker). Results indicated a high level of consistency for output among pairs of consecutive weeks, so Rambo et al. (1983) concluded that under certain circumstances, such as in a stable work environment, performance consistency may be obtained. However, the consistency did decrease as a function of time between the observation periods, suggesting that past performance may predict near-term rather than long-term performance. For example, how well an employee performed one week was more consistent with performance the week before than with performance in the previous year. However, the Rambo et al. (1983) study suffered from a small sample size and the failure to examine the effect of ability or learning on consistency. These two problems were addressed in Deadrick and Madigan's (1990) study.

Deadrick and Madigan (1990) studied sewing machine operators in a field study. They examined the effect of the time interval between occasions of performance measurement on performance consistency, the effect of experience and job aptitudes on performance consistency,



and the relationship between ability and performance. The approximately 900 sewing machine operators provided data on job performance and ability over a six-month period. Both objective performance measures such as the weekly production averages, monthly production averages, and quarterly production averages, and subjective measures such as a supervisory rating of production quantity and an overall performance rating were used to evaluate individuals' performance. Deadrick and Madigan (1990) found that the relative performance of the sewing machine operators was not stable over time, but declined as a function of the time interval between the occasions of performance measurement. These results were the same regardless of aptitude characteristics or experience. That is, performance consistency did decrease over time, and was not affected by experience and aptitudes. The authors concluded that skilled performance is unstable over time.

Hofmann et al. (1992) found direct support for this simplex effect when they analyzed performance data for baseball players (pitchers and batters) over a 10 to 30 year period. Measures of yearly batting average and yearly earned run average were chosen as indices of performance, and obtained from a baseball encyclopedia. They found that individuals did change their rank order over time, though the simplex pattern was observed only for pitchers, not batters.

Landis (2001) conducted a longitudinal study to evaluate the stability of team performance, focusing on 23 professional basketball teams. The measure of performance or productivity was each team's regular winning percentage over 10 years, from the 1984/1985 seasons through to the 1993/1994 seasons. Data was derived from a basketball encyclopedia. Landis (2001) found support for the simplex pattern as well; that is, teams who performed best at the first time period of performance were performing the worst at the last time period of

performance, and vice versa. Future research should continue to examine performance change in teams, as there is an increasing trend towards accomplishing critical projects using teams in organizations.

Howard and Jacobs (2000) used results from a computer-based task to measure performance over time. General cognitive ability and task-specific ability were hypothesized to play a role in initial and final performances. The participant sample consisted of 170 university students, who were required to complete a Tower of Hanoi task. The two performance indicators in this study were the number of moves required to complete the problem and the elapsed time. Howard and Jacobs (2000) found that individuals with high general ability performed better initially than individuals with low general ability, cognitive ability determined initial performance more than it did final performance, and task-specific ability was related to higher levels of final performance. Also, cognitive ability was found to be more influential than task-specific ability when predicting initial performance. Their findings supported the concept of dynamic criteria: When individuals were engaged in learning a task, their performance was slow and full of errors. Later, however, task-specific ability increased and performance improved. The fact that this study was completed in a lab and that performance was measured over a few hours pose some limitations, as there is no measure of continued performance over days, weeks, or months.

To address the question of how much change in performance can be attributed to test-retest unreliability in job performance, Sturman, Cheramie, and Cashen (2001) conducted a meta-analysis where only individual performance research studied over three or more time periods was included, to ensure that the data was longitudinal. They found that correlations of performance scores did decrease with an increase in time, but did not reach zero. Sturman et al.

also provide estimates of test-retest reliability and found that reliability was higher for subjective measures of individual job performance, rather than objective measures. The concept of dynamic criteria was supported. However, the meta-analysis is also valuable in that it differentiates between consistency, stability, and test-retest reliability, and then provides estimates of these three concepts (Sturman et al., 2001).

A nearly perfect simplex pattern was found by Ployhart and Hakel (1998), who examined the intraindividual variability of sales commissions for 303 securities analysts over two years (eight quarters). They also investigated the stability of predictors of sales performance (e.g., persuasion and empathy) over time. Sales performance was found to increase over time, but with individual differences in linear performance, indicating that not every individual increased in the same linear manner. Furthermore, those individuals who thought that they were perceived as empathetic had higher sales after one year, were more likely to increase their sales, and were less likely to experience a decrease in performance. This research supports the simplex pattern.

Although the research reviewed thus far provides evidence to support dynamic criteria via the simplex pattern, researchers need to think about better ways to investigate performance change; that is, the simplex pattern is not necessarily the most accurate method of revealing change over time. Several researchers have questioned whether the research thus far does indeed provide evidence for a simplex pattern, or whether the simplex pattern is simply an example of a statistical tool with a weak theoretical foundation (Ackerman, 1989; Barrett & Alexander, 1989). Several critics have raised questions about equating ability with initial performance (Ackerman, 1989), and have asserted that it is possible to use measures that may result in stable validity as the job skills are obtained and maintained. Initial performance does not correlate with later performance. However, Ackerman (1989) notes that this occurs for initial performance only.

Critics of dynamic criteria do not recommend the use of the simplex pattern and argue that the only inference that can clearly be made is that individuals do continuously change their rank order on performance with practice (Ackerman, 1989). The simplex process has been considered to be deficient in revealing anything about correlations between the predictor and criterion. In support of these views of simplex patterns, several studies have found that performance may be stable over time.

In a longitudinal study examining the performance of top-level executives, Russell (2001) investigated whether ratings based on prior life experiences predicted initial performance and change in performance. The participants in this study were 98 candidates considered for promotion as division general managers. Interviews were completed with all candidates, as well as ratings from 360-degree performance appraisals and assessment centers. Other criteria that were considered included objective measures such as profit and sales. Russell (2001) found that an overall rating predicted both the initial performance level and future performance trend for top-level executives across several criteria, such as superior ratings for annual fiscal and nonfiscal performance, amount of the annual management bonus, profit, and sales.

Another study that found support for stability of performance criteria studied rehabilitation counselors in order to generate stability coefficients for six performance criteria (Bolton, Neath, Bellini, & Cook, 1995). Using a sample of 50 rehabilitation services counselors, they examined six criteria of rehabilitation counselor performance, based on three statistical indexes: competitive closure rate, number of noncompetitive closures, and net case service encumbrance for noncompetitive closures. They found that performance indexes such as mean competitive closure rate did demonstrate stability over time, and that these performance criteria could be used for reliable counselor appraisal. However, these results were mixed, as for several

performance indexes (for clients with nonsevere disabilities), stability was a function of the length of the interval. Hence, these researchers found mixed results.

A framework proposed by Hanges, Schneider, and Niles (1990) looks at the stability of performance through an interactionist perspective; that is, behavior is a function of the interaction between a person and a situation. This interaction could help to explain conflicting findings thus far, as a simplex pattern may occur in different conditions (person-in-situation, situation alone, or person alone). Ratings were analyzed in three ways: across courses over time, a certain instructor in one particular course over time, and a certain instructor in all his/her courses over time. They hypothesized that stability would be highest when considering the ratings over time of an instructor teaching the same course. The stability of the ratings of the instructor alone or the course alone was hypothesized to yield lower stability coefficients. Data was obtained over six and a half years from 14,000 students in about 380 classes for approximately 300 faculty members every semester. The teaching evaluations contained a composite measure of teaching effectiveness as well as seven separate dimensions: organization, participation, availability, evaluation, personal style, learning, and recommendation. Studying the teaching ratings of these university instructors, Hanges et al. concluded that they had found “impressive” stability of performance over time, especially when the situation in which the participants were studied remained the same. Results did confirm their hypotheses; that is, the person-in-situation analysis did reveal higher stability than either the situation or person analyses. Interestingly, the results for the person analysis were found to contribute more to the stability in the person-in-situation analysis, as the situation analysis results were consistently weak.

To this point, results on the existence of dynamic criteria has been mixed. Although there have been several results that have supported the simplex pattern, and thus dynamic criteria, it is still not clear whether the simplex pattern is a useful tool, or whether dynamic criteria exist for all measures of performance. A more recent stream of research has approached the problem of dynamic criteria from an individual level of analysis, focusing on individual differences (Hoffman et al., 1992). Arguing that analyzing mean change in rank order can disguise systematic individual patterns of change, Hoffman et al. (1992) studied baseball players to determine whether systematic patterns existed. Hoffman et al. viewed performance over time as more of a developmental issue, and focused on individual trends rather than aggregate changes in performance. Using orthogonal polynomial growth curves to estimate the performance trend over time, they found support for their hypothesis that pattern of change was systematic for individuals, also replicating previous research using aggregated results. This analysis reveals more than just mean change in rank order, it actually discovers whether the changes in rank order were systematic at the individual level.

In another study, Hofmann, Jacobs, and Baratta (1993) stress a need to move beyond concluding that performance criteria are indeed dynamic to investigating the processes behind the change. They propose that dynamic criteria are synonymous with investigating individual change patterns, building on Hoffman et al.'s (1992) study. Using newer methodologies to study change, they first “statistically describe[d] the data for each individual and then look[ed] for between-person differences in the within-person parameters”. Using 319 newly-hired insurance sales personnel as a sample, Hoffman et al. (1993) gathered performance data (face value of insurance policies sold) each month for 36 months. Their three-year study, using orthogonal polynomial growth curves to study change, found that intraindividual change patterns were

indeed systematic (there was significant change over time), and that there were interindividual differences in intraindividual change (Hoffman et al., 1993).

Other researchers have also noted the importance of investigating intraindividual performance. Integrated latent growth curves and individual growth curves were used to answer the questions of whether performance is dynamic (Ployhart & Hakel, 1998), similar to Hoffman et al. (1992) and Hoffman et al. (1993). Ployhart and Hakel (1998) found that overall, criteria are dynamic over time and performance approximates a negatively accelerated learning curve; that is, there was not a simple linear relationship. Their findings suggest at least a degree of change in performance, as criterion intercorrelations decreased over time. These growth curves are argued to provide better and stronger evidence for dynamic criteria than the simplex pattern (Hofmann, Jacobs, & Gerras, 1992; Ployhart & Hakel, 1998).

### *Summary*

Is there a definitive answer as to whether individual level job performance is stable over time? Research to date has been inconclusive. Considering the amount of research that has found support for dynamic criteria, a tentative conclusion may be that performance is not stable, although it may be in certain contexts (Deadrick & Madigan, 1990; Hanges et al., 1990). Still, future research needs to continue to develop new approaches to assessing performance stability and it needs to address the methodological and conceptual flaws that were detailed in this review. Furthermore, some of the newer methodologies for modeling change in job performance (Hoffman et al., 1993; Ployhart & Hakel, 1998) should be researched further. Issues related to the unreliability of measures should also be addressed (Sturman et al., 2001).

Related to the issue of different approaches for modeling the stability of job performance, future research needs to examine the relationship between ability and performance where models allow for systematic change in ability level as learning increases (Henry & Hulin, 1990).

Furthermore, as more and more work is being done in teams, future research on teams should assess whether patterns regarding the stability of job performance hold for different types of teams.



## Conclusion

Understanding individual level job performance is vital to Canada's future competitiveness. As a nation we need to build a workforce that is talented, motivated, performs, and is productive. This literature review provides insight into some issues related to ensuring that the workforce performs while on the job. Although practitioners and researchers have been writing about the importance of job performance for almost a century, it is only during the last 30 years that researchers have theorized about it and tried to identify and understand issues related to its measurement. Fortunately, the body of literature that has emerged has answered a number of important questions about employee level job performance. However, future research needs to address unresolved issues and tackle new questions about this construct.

One significant achievement in the theorizing about job performance is the broadening of its scope to include positive behaviors that contribute to the social and psychological environment at work, named organizational citizenship behavior in this review. Twenty years of research have debated whether or not these behaviors should be incorporated into conceptualizations of job performance and if so how they should be defined. Researchers have made progress on this front and are moving toward a common definition and category scheme for describing this aspect of job performance. Researchers have also succeeded in incorporating the negative behaviors into theories of job performance. Counterproductive behaviors detract from the goals of the organization and pose problems for coworkers. Although a lot of research has tried to define this aspect of job performance, more work is needed before we have consensus on its definition and a category scheme that describes its structure.

Another significant milestone is the expansion of the measurement of employee level job performance to include ratings from organizational members who possess different perspectives

on employee performance. Traditionally, the supervisor or manager perspective was the only perspective that mattered. Today, researchers are examining the value of incorporating peer, self, and subordinate ratings into performance evaluations. Once again, researchers have responded to questions and concerns about the reliability, validity, and general use of 360-degree feedback. However, the research findings to date do not provide clear answers to the questions about whether these ratings are reliable, valid, or useful. Research findings are mixed. Although there is some evidence to suggest that ratings provided by members from different perspectives are reliable, predict promotions, and improve managerial effectiveness, other evidence points to the contrary. Furthermore, a number of methodological and conceptual limitations of the studies lead one to question the results of some of the research. Thus, there is a strong need for additional research that responds to these concerns.

Researchers have also made significant progress in determining whether employee level job performance is stable over time. Different approaches exist for addressing this issue. Research using one approach, which examines changes in an individual's rank ordering over time, has found support for a simplex pattern in the data indicating that individual level job performance is not stable but rather dynamic. However, current research is moving beyond this approach and looking at new methods for addressing this question. Future research needs to build on prior work and continue to be innovative in data analytic techniques.

## References

- Ackerman, P. L. (1989). Within-task intercorrelations of skilled performance: Implications for predicting individual differences? (A comment on Henry & Hulin, 1987). *Journal of Applied Psychology, 74*, 360-364.
- Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management, 27*, 479-495.
- Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49*, 141-168.
- Atwater, L.E., Ostroff, C., Yammarino, F.J., & Fleenor, J.W. (1998). Self-other agreement: Does it really matter? *Personnel Psychology, 51*, 577-598.
- Atwater, L., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology, 48*, 35-59.
- Atwater, L.E., Waldman, D.A., Atwater, D., & Cartier, P. (2000). An upward feedback field experiment: Supervisors' cynicism, reactions, and commitment to subordinates. *Personnel Psychology, 53*, 275-297.
- Atwater, L.E., & Yammarino, F.J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*, 141-164.
- Austin, J. T., Humphreys, L. G., & Hulin, C. L. (1989). Another view of dynamic criteria: A critical reanalysis of Barrett, Caldwell, and Alexander. *Personnel Psychology, 42*, 583-596.

Barrett, G. V., Caldwell, M.S., & Alexander, R.A. (1985). The concept of dynamic criteria: A critical reanalysis. *Personnel Psychology*, 38, 41-56.

Barrett, G. V., & Alexander, R. A. (1989). Rejoinder to Austin, Humphreys, and Hulin: Critical reanalysis of Barrett, Caldwell, and Alexander. *Personnel Psychology*, 42, 597-612.

Bolton, B., Neath, J., Bellini, J., Cook, D. (1995). Longitudinal stability and interrelationships of six criteria of rehabilitation counselor performance. *Rehabilitation Counseling Bulletin*, 39, 15-24.

Bommer, W.H., Johnson, J., Rich, G.A., Podsakoff, P.M., et al. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology*, 48, 587-605.

Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial/organizational psychology: Vol. 2.* (2nd ed., pp. 271-326). Palo Alto, CA: Consulting Psychologists Press.

Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance*, 6(1), 1-21.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W.C. Borman and Associates (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass Publishers.

Boswell, W.R., & Boudreau, J.W. (2000). Employee satisfaction with performance appraisals and appraisers: The role of perceived appraisal use. *Human Resource Development Quarterly*, 11, 283-299.

Bozeman, D.P. (1997). Interrater agreement in multi-source performance appraisal: A commentary. *Journal of Organizational Behavior*, 18, 313-316.

Bracken, D.W., Timmreck, C.W., & Church, A.H. (2001). History and development of multiple-source feedback as a methodology. In D.W. Bracken, C.W. Timmreck, & A.H. Church (Eds.), *The handbook of multisource feedback* (pp. 3-14). San Francisco: Jossey-Bass.

Brett, J.F., & Atwater, L.E. (2001). 360-degree feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, *86*, 930-942.

Brutus, S., London, M., & Martineau, J. (1999). The impact of 360-degree feedback on planning for career development. *Journal of Management Development*, *18*, 676-693.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette and L. M. Hough (Eds.), Handbook of industrial and organizational psychology (Vol. 1, p. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P. (1991). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial/organizational psychology: Vol. 2*. (2nd ed., pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W.C. Borman and Associates (Eds.), Personnel selection in organizations (pp. 35-70). San Francisco, CA: Jossey-Bass Publishers.

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

Cawley, B.D., Keeping, L.M., & Levy, P.E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, *83*, 615-633.

Church, A.H., & Bracken, D.W. (1997). Advancing the state of the art of 360-degree feedback. *Group and Organization Management*, 22, 149-161.

Church, A.H., Rogelberg, S.G., & Waclawski, J. (2000). Since when is no news good news? The relationship between performance and response rates in multirater feedback. *Personnel Psychology*, 53, 435-451.

Crino, M. D. (1994). Employee sabotage: A random or preventable phenomenon. Journal of Managerial Issues, 6(3), 311-330.

Deadrick, D. L., & Madigan, R. M. (1990). Dynamic criteria revisited: A longitudinal study of performance stability and predictive validity. *Personnel Psychology*, 43, 717-744.

Dominick, P.G., Reilly, R.R., & McGourty, J.W. (1997). The effects of peer feedback on team member behavior. *Group and Organization Management*, 22, 508-520.

Druskat, V.U., & Wolff, S.B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, 84, 58-74.

Edwards, M.R., Ewen, A.J., & Vendantam, K. (2001). How do users react to multisource feedback? In D.W. Bracken, C.W. Timmreck, & A.H. Church (Eds.), *The handbook of multisource feedback* (pp. 239-255). San Francisco: Jossey-Bass.

Farh, J.-L., Werbel, J.D., Bedeian, A.G. (1988). An empirical investigation of self-appraisal-based performance evaluation. *Personnel Psychology*, 41, 141-156.

Fedor, D.B., & Bettenhausen, K.L. (1989). The impact of purpose, participant preconceptions, and rating level on the acceptance of peer evaluations. *Group and Organization Studies*, 14, 182-197.

Fletcher, C., Baldry, C., & Cunningham-Snell, N. (1998). The psychometric properties of 360 degree feedback: An empirical study and a cautionary tale. *International Journal of Selection and Assessment*, 6, 19-34.

Folger, R. & Baron, R. A. (1996). Violence and hostility at work: A model of reactions to perceived injustice. In G. R. VandenBos & E. Q. Bulatao (Eds.), Violence on the job: Identifying Risks and Developing Solutions (pp. 51-85). Washington, LDC: American Psychological Association.

Frank, R. F. (1989). How passion pays: Finding opportunities in honesty. Business and Society Review, Summer, 20-28.

Funderburg, S.A., & Levy, P.E. (1997). The influence of individual and contextual variables on 360-degree feedback system attitudes. *Group and Organization Management*, 22, 210-235.

Furnham, A., & Stringfield, P. (1994). Congruence of self and subordinate ratings of managerial practices as a correlate of superior evaluation. *Journal of Occupational and Organizational Psychology*, 67, 57-67.

Furnham, A., & Stringfield, P. (1998). Congruence in job-performance ratings: A study of 360-degree feedback examining self, manager, peers, and consultant ratings. *Human Relations*, 51, 517-530.

George, J. M., & Brief, A. P. (1992). Feeling good-doing good: A conceptual analysis of the mood at work-organizational spontaneity relationship. Psychological Bulletin, 112(2), 310-329.

Ghiselli, E. E. (1956). Dimensional problems of criteria. *Journal of Applied Psychology*, 40, 1-4.

Ghiselli, E. E., & Haire, M. (1960). The validation of selection tests in the light of dynamic nature of criteria. *Personnel Psychology*, *13*, 225-231.

Giacalone, R. A., & Greenberg, J. (1997). Antisocial Behavior in Organizations. Thousand Oaks, CA: Sage.

Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, *75*, 561-568.

Greenberg, J., & Scott, K. S. (1996). Why do workers bite the hands that feed them? Employee theft as a social exchange process. In B. M. Staw & L.L. Cummings (Eds.), Research in Organizational Behavior (Vol. 18, pp. 111-156). Greenwich, CT: JAI.

Gruys, M. (1999). The dimensionality of deviant employee behavior in the workplace. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Halverson, S.K., Tonidandel, S., Barlow, C., & Dipboye, R.L. (2002). *Self-other agreement on a 360-degree leadership evaluation*. Poster session at the annual conference of the Society for Industrial and Organizational Psychology, Toronto, ON.

Hanges, P. J., Schneider, B., & Niles, K. (1990). Stability of performance: An interactionist perspective. *Journal of Applied Psychology*, *75*, 658-667.

Harris, M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, *41*, 43-61.

Hedge, J.W., & Teachout, M.S. (2000). Exploring the concept of acceptability as a criterion for evaluating performance measures. *Group and Organization Management*, *25*, 22-44.

Henry, R. A., & Hulin, C. L. (1987). Stability of skilled performance across time: Some generalizations and limitations on utilities. *Journal of Applied Psychology*, *72*, 457-462.



Henry, R. A., & Hulin, C. L. (1989). Changing validities: Ability-performance relations and utilities. *Journal of Applied Psychology, 74*, 365-367.

Hofmann, D. A., Jacobs, R., & Baratta, J. E. (1993). Dynamic criteria and the measurement of change. *Journal of Applied Psychology, 78*, 194-204.

Hofmann, D. A., Jacobs, R., & Gerras, S. J. (1992). Mapping individual performance over time. *Journal of Applied Psychology, 77*, 185-195.

Hollinger, R. C., & Clark, J. P. (1982). Formal and informal social controls of employee deviance. *Sociological Quarterly, 23*, 333-343.

Hollinger, R. C., & Clark, J. P. (1983). Theft by Employees. Lexington, MA: D. C. Heath.

Howard, M., & Jacobs, R. (2000). *Modeling performance over time*. Paper presented at the Fifteenth Annual Society for Industrial and Organizational Psychology Conference, New Orleans, LA. April 15, 2000.

Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin, 107*, 328-340.

Hunt, S. T. (1996). Generic work behavior: An investigation into the dimensions of entry-level, hourly job performance. *Personnel Psychology, (49)*, 51-83.

Jansen, P., & Vloeberghs, D. (1999). Multi-rater feedback methods: Personal and organizational implications. *Journal of Managerial Psychology, 14*, 455-476.

Johnson, J.W., & Ferstl, K.L. (1999). The effects of interrater and self-other agreement on performance improvement following upward feedback. *Personnel Psychology, 52*, 271-303.

Katz, D., & Kahn, R. L. (1978). The social psychology of organizations. New York: Wiley.

Keeping, L.M., & Levy, P.E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology, 85*, 708-723.

Landis, R. S. (2001). A note on the stability of team performance. *Journal of Applied Psychology, 86*, 446-450.

Latham, G.P., & Locke, E.A. (1991). Self regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50*, 212-247.

Latham, G.P., & Wexley, K.N. (1994). *Increasing productivity through performance appraisal* (2nd ed.). New York: Addison-Wesley.

London, M., Smither, J.W., & Adsit, D.J. (1997). Accountability: The Achilles' heel of multisource feedback. *Group and Organization Management, 22*, 162-184.

McEnery, J.M., & Blanchard, P.N. (1999). Validity of multiple ratings of business student performance in a management situation. *Human Resource Development Quarterly, 10*, 155-172.

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

Mount, M.K., Judge, T.A., Scullen, S.E., Sytsma, M.R., & Hezlett, S.A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology, 51*, 557-576.

Murphy, K. R. (1989). Dimensions of job performance. In Dillon R, Pellingrino J (Eds.), Testing: Applied and theoretical perspectives (p. 218-247). New York: Praeger.

Murphy, K. R. (1993). Honesty in the Workplace. Belmont, CA: Brooks/Cole.

Murphy, K.R., Cleveland, J.N., & Mohler, C.J. (2001). Reliability, validity, and meaningfulness of multisource ratings. In D.W. Bracken, C.W. Timmreck, & A.H. Church

(Eds.), *The handbook of multisource feedback* (pp. 130-148). San Francisco: Jossey-Bass.

Nease, A.A., Mudgett, B.O., & Quinones, M.A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology, 84*, 806-814.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78(4)*, 679-703.

Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance, 10(2)*, 85-97.

Organ, D. W. (1988). Organizational citizenship behavior: The good soldier syndrome. Lexington, MA: Lexington Books.

Ployhart, R. E., & Hakel, M. D. (1998). The substantive nature of performance variability: Predicting interindividual differences in intraindividual performance. *Personnel Psychology, 51*, 859-901.

Raelin, J. A. (1994). Three scales of professional deviance within organizations. *Journal of Organizational Behavior, 15*, 483-501.

Rambo, W. W., Chomiak, A. M., & Price, J. M. (1983). Consistency of performance under stable conditions of work. *Journal of Applied Psychology, 68*, 78-87.

Reilly, R.R., Smither, J.W., & Vasilopoulos, N.L. (1996). A longitudinal study of upward feedback. *Personnel Psychology, 49*, 599-612.

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal, 38(2)*, 555-572.

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. Academy of Management Journal, 38(2), 555-572.

Robinson, S. L., & Greenberg, J. (1998). Employees behaving badly: Dimensions, determinants, and dilemmas in the study of workplace deviance. In C. L. Cooper and D. M. Rousseau (Eds.) Trends in Organizational Behavior: Vol. 5 (pp. 1-30). Chichester, West Sussex: John Wiley & Sons.

Rothe, H. R. (1978). Output rates among industrial employees. *Journal of Applied Psychology*, 63, 40-46.

Rotundo, M. (2000). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. Dissertation Abstracts International, 61(5-A), 2058 (UMI No. AAI9973011).

Rotundo, M. & Sackett, P.R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy capturing approach. *Journal of Applied Psychology*, 87(1), 66-80.

Russell, C. J. (2001). A longitudinal study of top-level executive performance. *Journal of Applied Psychology*, 86, 560-573.

Sala, F., & Dwight, S. (2002, April). *Predicting executive performance with multi-rater surveys: Who you ask matters*. Poster session at the annual conference of the Society for Industrial and Organizational Psychology, Toronto, ON.

Schaubroeck, J., & Green, S.G. (1989). Confirmatory factor analytic procedures for assessing change during organizational entry. *Journal of Applied Psychology*, 74, 892-900.

Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage Publications.

Skarlicki, D. P., & Folger, R. (1997). Retaliation in the workplace: The roles of distributive, procedural, and interactional justice. Journal of Applied Psychology, *82*, 416-425.

Smith, P. C. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 745-775). Chicago, Illinois: Rand McNally College Publishing Co.

Smith, E.M., & Helland, K.R. (2002). *Self-other rating agreement in multi-source feedback: Antecedents and correlates*. Poster session at the annual conference of the Society for Industrial and Organizational Psychology, Toronto, ON.

Smither, J.W., London, M., Reilly, R.R., Flautt, R., Vargas, Y., & Kucine, I. (2002). *Does discussing multisource feedback with raters enhance performance improvement?* Poster session at the annual conference of the Society for Industrial and Organizational Psychology, Toronto, ON.

Smither, J.W., London, M., Vasilopoulos, N.L., Reilly, R.R., Millsap, R.E., & Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology*, *48*, 1-34.

Smither, J.W., & Walker, A.G. (2001). Measuring the impact of multisource feedback. In D.W. Bracken, C.W. Timmreck, & A.H. Church (Eds.), *The handbook of multisource feedback* (pp. 256-271). San Francisco: Jossey-Bass.

Sturman, M. C., Cheramie, R. A., & Cashen, L. H. (2001). *The consistency, stability, and test-retest reliability of employee job performance: A meta-analytic review of longitudinal findings*. Working Paper Series: The Center for Hospitality Research.

Van Dyne, L., Cummings, L. L., & Parks, J. M. (1995). Extra-role behaviors: In pursuit of construct and definitional clarity (A bridge over muddied waters). In B. M. Staw & L. L.

Cummings (Eds.), Research in organizational behavior (Vol. 17, pp. 215-285). Greenwich, CT: JAI Press.

Van Velsor, E., Taylor, S., Leslie, J. (1993). An examination of the relationship among self-perception accuracy, self-awareness, gender, and leader effectiveness. *Human Resource Management, 32*, 249-264.

Viswesvaran, C., Ones, D.S., Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574.

Waldman, D.A., Atwater, L.E., & Antonioni, D. (1998). Has 360 degree feedback gone amok? *Academy of Management Executive, 12*, 86-93.

Walker, A.G., & Smither, J.W. (1999). A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology, 52*, 393-423.

Wood, R.E., Allan, J., Pillinger, T., & Kohn, N. (2000). 360 feedback: Theory, research and practice. In T. Travaglione & V. Marshall (Eds.), *Human resource management theory and practice* (pp. 209-228). Sydney: McGraw Hill.

Table 1

A Summary of Various Efforts to Describe the Domain of Job Performance.

Author	Components	Description
Murphy (1989)	Task performance. Interpersonal relations. Destructive or hazardous behaviors. Down-time behaviors.	Accomplishment of duties and responsibilities. Cooperating; communicating; exchanging job-related information. Violating security and safety; destroying equipment, accidents. Substance abuse; illegal activities.
Campbell (1990)	Job-specific task proficiency. Non-job-specific task proficiency. Written and oral communication proficiency. Demonstrating effort. Maintaining personal discipline. Facilitating peer and team performance. Supervision and leadership. Management and administration.	Core technical tasks. Tasks not specific to a given job. Preparing written materials or giving oral presentations. Exerting extra effort; willing to work under adverse conditions. Avoid negative or adverse behaviors (e.g., substance abuse). Support and assist peers; reinforce participation. Influence; setting goals; rewarding and punishing. Organize people and resources; monitor progress; problem-solve.
Borman & Brush (1993)	Technical activities. Leadership and supervision. Interpersonal dealings. Useful personal behavior.	Planning; demonstrating technical proficiency; administration. Guiding; directing; motivating; coordinating. Communicating; maintaining a good organizational image and working relationships. Working within the guidelines and boundaries of the organization.
Hunt (1996)	Generic Work Behaviors	Adherence to confrontational rules Industriousness Thoroughness Schedule flexibility Attendance Off-task behavior Unruliness Theft Drug misuse

[Adapted from Rotundo, M. & Sackett, P.R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy capturing approach. *Journal of Applied Psychology*, 87(1), 66-80.]

Table 2

A Summary of Various Efforts to Conceptualize Task, Organizational Citizenship and Counterproductive Job Performance

Author	Component	Behavioral Categories
Brief & Motowidlo (1986)	Prosocial Organizational Behavior	Assisting co-workers with job-related matters. Showing leniency. Providing services/products to consumers in organizationally consistent ways. Providing services/products to consumers in organizationally inconsistent ways. Helping consumers with personal matters unrelated to organizational services/products. Complying with organizational values, policies, and regulations. Suggesting procedural, administrative, or organizational improvements. Objecting to improper directives, procedures, or policies. Putting forth extra effort on the job. Volunteering for additional assignments. Staying with the organization despite temporary hardships. Representing the organization favorably. Assisting co-workers with personal matters.
Organ (1988)	Organizational Citizenship Behavior	Altruism Conscientiousness Sportsmanship Courtesy Civic Virtue
George & Brief (1992)	Organizational Spontaneity	Helping co-workers. Protecting the organization. Making constructive suggestions. Developing oneself. Spreading goodwill.
Borman & Motowidlo (1993)	Task Performance Contextual Performance	Formally recognized as part of the job and contribute to the organization's technical core. Discretionary; not necessarily role prescribed; contribute to social/psychological environment.
Raelin (1994)	Professional Deviant/ Adaptive	Work-scale (e.g., unethical practices, absenteeism, work-to-rule, bootlegging) Self-scale (e.g., flaunting of external offers, rationalization, alienation, apathy) Career-scale (e.g., premature external search, external performance emphasis)
Van Dyne, Cummings, & Parks (1995)	Extra-Role Behavior	Affiliative/Promotive (e.g., helping and cooperative behaviors) Challenging/Promotive (e.g., constructive expression of challenge) Challenging/Prohibitive (e.g., criticism of situation to stop inappropriate behavior) Affiliative/Prohibitive (e.g., unequal power or authority)
Robinson & Bennett (1995)	Employee Deviance	Property deviance Production deviance Political deviance Personal aggression

[Adapted from Rotundo, M. & Sackett, P.R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy capturing approach. *Journal of Applied Psychology*, 87(1), 66-80.]