

Reactions to Different Types of Forced Distribution Performance Evaluation Systems

Brian D. Blume · Timothy T. Baldwin ·
Robert S. Rubin

Published online: 5 March 2009
© Springer Science+Business Media, LLC 2009

Abstract

Purpose We isolate and describe four key elements that distinguish different forms of forced distribution systems (FDS). These key elements are the consequences for low performers, differentiation of rewards for top performers, frequency of feedback, and comparison group size. We examine how these elements influence respondents' attraction to FDS.

Design/methodology/approach Undergraduate students ($n = 163$) completed a policy capturing study designed to determine how these four FDS elements influence their attraction to FDS. We examine the relative importance of these elements that most influence attraction to different FDS, as well as individual attributes (i.e., cognitive ability, gender, and major) that may affect those preferences.

Findings Respondents were most attracted to systems with less stringent treatment of low performers, high differentiation of rewards, frequent feedback and large comparison groups. Consequences for low performers were nearly twice as influential as any other element.

Respondents with higher cognitive ability favored high reward differentiation and males were less affected by stringent consequences for low performers.

Implications Before practitioners implement FDS, it would be prudent to consider all four elements examined in this study—with the treatment of low performers being the most salient issue. Future accounts of FDS should clarify the nature of these elements when reporting on FDS. Such precision will be useful in generating a knowledge base on FDS.

Originality/value We add precision to the discussion of FDS by identifying four key elements. This is one of the first studies to examine perceptions of FDS from a rater perspective.

Keywords Forced distribution · Performance management · Performance evaluation · Policy capturing · Relative performance appraisal · Force ranking

Introduction

Performance evaluation systems are one of the most pervasive and important human resources systems in organizations today (Murphy and Cleveland 1995; Judge and Ferris 1993). Despite their ubiquitous use, previous research has also documented significant shortcomings in the application of performance evaluations, including many forms of biases stemming from rating errors, sources of performance information and individual differences (Arvey and Murphy 1998). Of these various shortcomings, one of the most common forms of performance rating biases is the tendency on the part of raters to provide lenient or inflated ratings (Bretz et al. 1992; Rynes et al. 2002). This systematic bias usually results in a lack of differentiation

Received and reviewed by former editor, George Neuman.

B. D. Blume (✉)
School of Management, University of Michigan,
3119 WSW Bldg., Flint, MI 48502, USA
e-mail: blume@umflint.edu

T. T. Baldwin
Kelley School of Business, Indiana University,
Bloomington, IN, USA
e-mail: baldwint@indiana.edu

R. S. Rubin
Kellstadt Graduate School of Business, DePaul University,
Chicago, IL, USA
e-mail: rrubin@depaul.edu

between high and low performers, yielding inaccurate performance information (Guralnik et al. 2004; Jawahar and Williams 1997). For both administrative and developmental performance evaluations, this lack of differentiation on the part of raters can be problematic, leaving organizations with little variation of inputs when making important personnel decisions such as promotions, terminations or training opportunities.

In recognition of this systematic bias, researchers over the past two decades have explored methods that might increase differentiation and accuracy (Goffin et al. 1996). The focus of this research has primarily highlighted the effects of two general rating systems: absolute and relative evaluations. In absolute rating systems, individual performance is assessed against a particular standard, whereas in relative systems individual performance is determined by comparing people against one another (Duffy and Webber 1974). Although there are advantages to both types, in terms of improving performance differentiation on a variety of criteria, a few studies have indicated that relative rating systems may be more effective than absolute evaluations (Heneman 1986; Nathan and Alexander 1988; Wagner and Goffin 1997). Given these notable advantages to relative systems, we would expect that relative performance appraisal systems (e.g., relative percentile method, ranking, etc.) would be implemented more often than they are currently in organizations. One reason that organizations may avoid using relative systems is that research suggests that ratees hold rather negative perceptions of relative systems (Roch et al. 2007).

Recently, however, there has been a revival of sorts regarding the usefulness of relative systems taking place in organizations in the form of forced distribution systems (FDS). FDS are one form of relative performance evaluation that were developed in an attempt to deal directly with the problems of rater leniency and the lack of differentiation of performance evaluation ratings (McBriarty 1988). FDS do so by “forcing” managers to discriminate between high and low performers. FDS generally involves either sorting employees into predetermined performance categories using a defined distribution curve (i.e., a set percentage of high, average and low performers) or ranking them on the basis of relative performance (Guralnik et al. 2004). Despite limited evidence of its overall practical value to organizations, the use of FDS proliferated greatly (Pfeffer and Sutton 2006). Recent surveys suggest that at least 20% of American businesses now use FDS, including many admired and progressive organizations such as Heinz, Microsoft, American Express, and Goldman Sachs (Bates 2003; Olson and Davis 2003).

Perhaps no other figure has contributed more to the increased use of FDS than high-profile former General Electric (GE) executive Jack Welch (Bossidy and Charan

2002; Tichy and Sherman 2001). Welch has extolled FDS as being an efficient and pragmatic means of “rewarding doers” and “building muscle” for the organization. These and other authors often point to the success of GE in developing executive talent as evidence of the efficacy of FDS. Indeed, FDS at GE and other organizations are widely viewed as more than just a means of evaluating performance; but rather, central to the development and succession planning processes and is thought to be the cornerstone of achieving a performance-oriented culture.

Forced distribution, however, is not without its articulate critics. Well-known authors like Jeffrey Pfeffer and Malcolm Gladwell, condemn FDS as dysfunctional and suggest that such systems are hazardous to an organization’s culture and performance (Pfeffer and Sutton 2006). Critics often point to examples from organizations such as Ford Motor Company (Colvin 2001; Shirouzu 2001), which had a well-publicized *unsuccessful* experience with FDS (Gladwell 2002; Pfeffer 2001). Ford’s corporate culture did not appear to support FDS, as many employees who had gotten positive feedback for years were suddenly told that they were now underperformers (Colvin 2001). Dozens of Ford employees and ex-employees sued the company over the program (Colvin 2001). Among those critical of forced distribution practices, some have a philosophical objection to the concept of forced distribution in general, while others simply take issue with the “way it is often done.” Despite impassioned anecdotal accounts on both sides of the debate, little empirical research has emerged. FDS have generally been discussed as one-dimensional phenomenon and few distinctions of critical FDS elements, or the impact of these different elements on perceptions and outcomes, have been investigated.

With the above issues in mind, the purpose of the present paper is twofold. First, we contribute to the theoretical literature on FDS by isolating and clarifying the fundamental elements of different manifestations of FDS in organizations. Recognizing that not all FDS are alike in design or practice our goal was to identify those elements that are most salient and likely to influence perceptions and outcomes. To do so, we reviewed the existing literature on FDS, looking for those system elements that have been most frequently referenced or debated and raise important, but unanswered questions. Second, using the key system elements, we conduct a policy capturing study designed to determine how FDS elements influence the attraction to a FDS. In addition, we examine the relative importance of these elements that most influence attraction (or aversion) to different FDS, as well as individual attributes that might materially affect those preferences. As such, we build on the burgeoning literature regarding perceptions of performance appraisal systems (Levy and Williams 2004). These perceptions of performance management systems have been shown to be

related to peoples' engagement and satisfaction with a given system (Wright 2002; Mount 1984). As Cawley et al. (1998, p. 616) point out, "After all, one may develop the most technically sophisticated, accurate appraisal system, but if that system is not accepted and supported by employees, its effectiveness ultimately will be limited."

Hypothesis Development

One of the intriguing potential effects of FDS concerns their impact on an organization's labor market (Scullen et al. 2005). As noted, FDS have been promoted on the basis of attracting and retaining more talented employees despite a paucity of research demonstrating FDS's actual impact on attraction and retention. Indeed, a performance management system and its subsequent rewards is one important and legitimate way in which an organization can differentiate itself (Gerhart and Milkovich 1990) when looking to attract and retain talent.

Research on personnel recruiting has found that job seekers consider human resource (HR) systems when they develop beliefs about an organization's culture and consider whether they want to work for a company (Breugh and Starke 2000; Cable and Judge 1996). With respect to FDS, Scullen et al. (2005, p. 28) observed, "...if job seekers become aware of a company's FDRS and consider it too stressful or risky, they might not apply. ...It is certainly possible, however, that other high-quality applicants would see such a system as one where their contributions would be recognized and rewarded. These people might be eager to work in this type of environment." Similarly, Bretz and Judge (1994) noted, "reward system characteristics reflect fundamental differences in what the organization deems valuable." Further, Judge and Bretz (1992) found that organizational values were an important determinant of job choices and that individuals preferred jobs in organizations which displayed value preferences similar to their own. In addition, there is an increasing recognition in the performance appraisal literature that ratee reactions to appraisals are important (Hedge and Teachout 2000; Keeping and Levy 2000). When put simply, perceptions of FDS are likely to differ, and these perceptions will influence relative attraction or aversion to an organization using FDS.

One particularly salient issue associated with attraction to FDS is the perception of fairness with respect to both the process used and derivation of subsequent rewards. Prior research has shown that employee perceptions of fairness can have important effects on the outcomes attained by performance evaluation systems (Taylor et al. 1998; Folger and Konovsky 1989). Indeed, perceptions of fairness tend to be strongly correlated with negative reactions and withdrawal of individuals within an organization (Korsgaard

and Roberson 1995; Colquitt et al. 2001). Thus, when considering FDS, it seems likely that fairness or justice perceptions will be critical influences on whether such systems have a chance for success. Prior research (Blume et al. 2005) has found that individuals do differ significantly in their perceived fairness of FDS and ultimately their attraction to organizations using FDS. Thus, perceptions of fairness associated with FDS may significantly influence attraction to such a system.

Drawing upon models of justice perceptions as our theoretical underpinning, we sought to isolate the most critical FDS design elements that might be expected to relate to attraction to such a system. Although there has been no prior taxonomic work of which we are aware, our review of the research and popular literature revealed four recurring FDS elements which seem most likely to induce fairness perceptions and influence attraction to FDS: (a) the consequences for low performers (e.g., termination vs. development), (b) the differentiation of rewards among high and lower performers, (c) comparison group size, and (d) the frequency and consistency of feedback. Although these elements may be found in other performance management systems, FDS and other relative performance appraisal systems highlight the importance of each of these elements. We introduce these four elements below and hypothesize how each of them will influence the attraction to FDS.

Consequences for Low Performers

A ubiquitous issue in discussions of FDS is what to do with those rated on the low end of the scale. Welch argues that low performers must not receive anything in the way of rewards and that removing the bottom 10% is essential to FDS (Welch 2001). Levinson (2003) similarly suggests that poor performers should generally be terminated or at the very least be given a warning. These authors and others contend that development efforts for these "C players" are counter-productive and that a better solution is to reserve development efforts for A's and remove C players from their jobs (Grote 2002).

It could also be that employees may approve of an organization's decision to eliminate underperformers. Axelrod et al. (2002) reported that, of thousands of senior managers they polled, "96% of them said they would be delighted if their companies moved more aggressively on low performers." Equity theory indicates that employees compare themselves to each other in terms of inputs and outcomes (Walster et al. 1978). As Scullen et al. (2005) point out, if low performers who do not contribute are removed from the organization, high performers might feel that an equitable balance is being established and might be more motivated to continue their high quality work and remain with the organization.

In practice, however, reports from several firms (e.g., Ford, Goodyear) that have had well documented, unsuccessful experiences with FDS suggest that the labeling and dismissal of C performers were frequently viewed as unfair and inequitable and perhaps the most morale-damaging element. Organ (1990) states that having a fundamental respect for human dignity is a critical factor in perceptions of fairness and that “even the most incompetent and incorrigible subordinate has the right to be treated civilly.” Indeed, in a number of cases, firms have started by asking those identified as C’s to leave, but then later softened their stance toward C’s and offered more development and training opportunities with a goal of helping them to improve performance (Bates 2003; Shirouzu 2001). Resistance to firing of C players may stem in part from the difficulty in establishing clear and transparent performance criteria for many job roles in real organizational contexts and in part from the reality that attribution biases generally preclude most people from viewing themselves as a C in any context. Moreover, there is a general humanistic concern with any pejorative labeling of people as C players and firing a certain number of such people may cause wholesale rejection of a system that might otherwise have been supported.

In sum, it seems likely that the consequences for low performers will be a salient issue related to the attractiveness of FDS. Although some prior evidence suggests that at least some people will see removal of the lowest-rated employees as desirable, accounts from the field suggest that a less stringent, more developmentally oriented treatment of such low performers would be most attractive.

Hypothesis 1 Stringent consequences for poor performers are negatively associated with attraction to FDS.

Differentiation of Rewards

Perhaps the most commonly advocated advantage of forced distribution is that the method helps build a high-performance, merit-based culture by ensuring that managers better differentiate among high, average, and low performers. Equity theory suggests that in these situations the most productive performers are likely to perceive inequity if they receive similar rewards as the majority of others (Adams 1965).

FDS protocol generally does prescribe significant performance distinctions. For example, Welch has argued that top performers should be getting raises that are two to three times the size given to the next level of performers (Welch 2001). At the same time, there has been a great deal of recent attention regarding wage dispersion or the ratio of the highest paid employee to the lowest paid (Bloom 1999; Pfeffer 1998; Pfeffer and Sutton 2006). This research has demonstrated that wide disparities, particularly among

people with the same or similar job requirements, are viewed as being unfair and inequitable (Gerhart and Milkovich 1992). Some have suggested that such material reward differences among people in comparable jobs at comparable levels have a deleterious effect on teamwork and cooperative behavior (Lawler 2003; Pfeffer 2001). Prior research has clearly demonstrated that people do pay a great deal of attention to reward differentiation and are often suspect of the legitimacy of those differences (Kanfer 1990). We therefore hypothesize that *very* high levels of reward differentiation will be generally viewed as unfavorable.

Hypothesis 2 Higher levels of reward differentiation are negatively associated with attraction to FDS.

Comparison Group Size

One of the underlying premises of a FDS is the well-documented phenomenon of a normal distribution, commonly known as the “bell-curve.” The notion is that when measured in large enough samples, most data (in this case, performance ratings) will distribute predictably in accord with the normal curve. However, among the most common laments of FDS is that raters must label some as ‘low performers’ even in a small peer group where everyone is performing well in an objective or absolute sense (Gary 2001).

Here again, however, staunch advocates of FDS are unbending. Welch, for example, is known for advocating what he labels the “vitality curve” (i.e., 20% A’s, 70% B’s and 10% C’s) in all units regardless of size. Indeed, one of the more legendary stories is of an informal discussion between Welch and a manager in a New York retail store. The store manager explained that he had 20 people in his sales force and asked Welch whether he really had to let two go? Welch replied, “Yes” (Colvin 2001).

Consistent with the discussion above and justice perceptions, the salient issue is whether the size of a particular group would be perceived to be large enough to warrant fair comparisons (Lawler 2002). Of course, all things equal, ratees and FDS advocates alike would probably prefer larger comparison groups. Thus, at one level the hypothesis below may seem self-evident. However, a key issue in exploring comparison group size was to determine the *relative* importance of that variable in relation to other key elements of FDS.

Hypothesis 3 Comparison group size is positively associated with attraction to FDS.

Frequency and Consistency of Feedback

Studies of performance evaluation systems and their effects on attitudes have found that one of the key factors in

perceptions of fairness is feedback (Landy et al. 1978; Organ 1990). For example, Landy et al. (1978) found that frequency of evaluation was significantly related to perceptions of fairness and accuracy of performance evaluation. Employees' sense of fairness is also likely violated when they are given inconsistent feedback. Where FDS have been implemented, some employees have claimed that they had always received good performance evaluations but then suddenly became a non-performer or C-player (Bates 2003). For instance, before Ford implemented its forced ranking system, 98% of its management employees were routinely ranked as fully meeting expectations under its former appraisal system (Olson and Davis 2003).

Welch (2001) suggests that one of the key reasons a FDS is effective at GE is the performance culture supported by candid feedback at every level. The frequency of formal feedback given to employees would be expected to be especially important in an organization where FDS requires differentiation among employees and where there are high stakes for the outcomes of manager rankings. Therefore, employees should see a FDS as being fairer if surprises are avoided via frequent and consistent feedback. This would be expected despite the fact that some feedback may negatively affect performance (Kruger and DeNisi 1996).

Although the benefits of frequent feedback have been well-documented, it does demand an exceptional commitment of management time and expertise. Further, as the size of the comparison group increases beyond the span of control of individual managers, it becomes increasingly difficult and complex to create a context where all people in a comparison group are given frequent, consistent, and meaningful feedback that would preclude surprises. Therefore, if feedback frequency and consistency is not of relatively high importance to the perceived attractiveness of FDS, then other issues may rightly assume more attention in the design and execution of such systems.

In any case, drawing on our conceptual framework of justice, it is expected that the frequency and consistency of feedback will be among the most salient issues to individuals rated under FDS. It is likely that individuals will perceive FDS as more procedurally just if they believe they will receive frequent feedback and have the opportunity to improve their performance. We therefore hypothesize:

Hypothesis 4 Frequent performance feedback is positively associated with attraction to FDS.

As previously noted, in addition to the above hypotheses, we were also interested in the *relative* importance of each FDS element in forming overall perceptions of system attractiveness. That is, we wanted to know which elements have the most important influence on the attractiveness of FDS and thus we designed the study and analyses accordingly.

As a final exploratory part of the study, we examine how certain individual differences may influence respondents' weights of these FDS elements. Blume et al. (2005) found that male graduates and those with high cognitive ability were more attracted to FDS than females or those with lower cognitive ability. Therefore, we included cognitive ability and gender as variables in the study. Since individuals with high cognitive ability are likely to be higher performers than those with lower cognitive abilities (Schmidt and Hunter 1998), this group of individuals is particularly salient with regard to developing performance management systems. That is, given that one reason organizations may implement FDS is to attract, develop and retain high performers, understanding how those with high cognitive ability may react to specific FDS elements is critically important to developing a system that supports high performance. Further, concerns with diversity and the ability to attract top females makes any differential impact of FDS elements based on gender of interest as well. Finally, it is possible that there could be systematic effects based on individuals' exposure to and knowledge of various organizations based upon their educational background. Since participants of the study were from business and non-business backgrounds, we also included business major as a dichotomous variable in the study.

Method

Participants were 163 primarily upper-level, undergraduate students (i.e., 91% were in their junior or senior year) enrolled in a management course at a large Midwestern university. These students were nearly all traditional students between the ages of 19 and 24. The sample was 68% male and 71% business majors. Non-business majors (e.g., liberal arts majors) accounted for the remaining 29% of the sample.

Policy Capturing Procedure

This study employed a policy capturing approach.¹ A key methodological advantage of policy capturing is that it allows for systematic and controlled manipulation and sampling of independent variables (Aiman-Smith et al.

¹ Policy capturing uses regression techniques to capture the cognitive processes underlying judgments. The method has been used to study variety of decision-making processes within organizations (see Karren and Barringer 2002 for a listing of policy capturing studies appearing in top-tier journals), including organizational attraction, job search, and job termination decisions (e.g. Aiman-Smith et al. 2001; Cable and Judge 1994; Rousseau and Anton 1988). For more information on policy capturing, see Karren and Barringer (2002) or Aiman-Smith et al. (2002).

2002). Policy capturing is an alternative to the direct estimation techniques (e.g., self-report), which give little indication of how rankings are used in actual decision making, demand greater self-insight than is likely to be possessed by decision makers, and are frequently criticized for eliciting responses subject to social desirability (Jur-gensen 1978; Schwab et al. 1987). Policy capturing alleviates some of these issues because individuals are placed more fully into the decision-making role, where subjects evaluate attributes of organizations rather than directly state preferences for specific organizational attributes (Karren and Barringer 2002).

In the present study we utilized policy capturing to examine the relative importance of consequences of poor performance, reward differentiation, comparison group size, and frequency of feedback on perceived attraction to FDS. Participants read scenarios that included one of the two levels for each of these four elements (see Table 1). After reading the scenarios, participants indicated how attracted they would be to a company that uses the FDS described in the scenario. After completing all scenarios, participants completed questions concerning demographics and individual difference variables.

Table 1 Forced distribution system elements

Consequences of poor performance	
Lower:	The 10% of employees receiving the lowest ‘C’ ranking are not given pay increases or bonuses. They receive additional training and coaching and are not usually terminated
Higher:	The 10% of employees receiving the lowest ‘C’ ranking are not given pay increases or bonuses. If they do not improve their performance, they usually either resign or are terminated
Reward differentiation	
Lower:	This company uses these rankings to distribute rewards to 20% of the top A-performers that are 1–2 times more than to the B-performers. These rewards include pay increases, company stock options and bonuses. One example would be that A-performers might receive a bonus of \$4,000 while the B-performers would receive a bonus of \$2,000
Higher:	This company uses these rankings to distribute rewards to 20% of the top A-performers that are 3–4 times more than to the B-performers. These rewards include pay increases, company stock options and bonuses. An example would be that A-performers might receive a bonus of \$8,000 while B-performers would receive a bonus of \$2,000
Comparison group size	
Lower:	You can expect to be compared to and ranked in a group of 10 of your peers
Higher:	You can expect to be compared to and ranked in a group of 50 of your peers
Frequency of feedback	
Lower:	You can expect that your supervisor will give you formal feedback about your performance annually
Higher:	You can expect that your supervisor will give you formal feedback about your performance at least 3–4 times a year

Effective policy capturing design requires enough scenarios and factors to yield stable estimates, but not so many that respondents become bored or fatigued. Our design is in line with the recommendations of Aiman-Smith et al. (2002) and Karren and Barringer (2002). By completely crossing these factors, 16 discrete scenarios (i.e., 2^4) were created. Full-factorial orthogonal designs yield the most stable and unambiguous estimates (Karren and Barringer 2002) and permits the assessment of how much weight each factor carried with the group of respondents. Four replicated scenarios were also included to assess within-rater judgment consistency, bringing the total number of scenarios to 20.

Materials

Respondents were instructed to read the descriptions of each of the companies and to indicate how attractive it would be to work for that company. They were told to assume that they are nearing the completion of their degree and are looking for a job. They were also instructed to assume that the characteristics of the companies (e.g., industry, size, type of position, location, salary offer, amount of money allocated for raises, etc.) were similar to one another and to other job offers they might expect to receive. Consistent with the definition given earlier, a FDS was described to participants in the following way:

Each of the companies uses a performance management system in which employees are ranked against a peer group. Managers assign each employee to one of 3 categories, with 20% of employees receiving the top ‘A’ ranking, 70% receiving the middle ‘B’ ranking, and 10% receiving the bottom ‘C’ ranking. You would receive one of these rankings every year”.

Factor Levels

We use two levels for each of the four elements of the FDS (Aiman-Smith et al. 2002), and each of these levels are listed in Table 1. Given the limited amount of empirical research on FDS elements, we drew on existing anecdotal accounts from practitioner-oriented publications as well as relied on our collective experience to determine the factor level ranges, which are discussed further below.

Consequences of Poor Performance

Given the ongoing debate over whether to ask the bottom-ranked individuals to leave or not, we decided to make this the focus of this factor. Therefore, in the lower (or less stringent) condition for consequences for poor performance, we suggested that those individuals ranked as C’s

do not receive pay increases or bonuses, but do receive additional training and are not usually terminated. In the higher (or more stringent) condition, we state that individuals ranked as C's not only are not given pay increases or bonuses, but also either usually resign or are terminated if they do not improve their performance. The more stringent condition was worded in this way because, to the authors' knowledge, all advocates of removing the C-players prescribe to carrying this out in a humane way (e.g., giving these employees the opportunity to resign or change positions within the company, offering severance packages). The main point is that those ranked as C's will most likely have to leave the company if their performance does not improve.

Differentiation of Rewards

Welch (2001) believes that differentiations made by FDS must be supported by the reward system (e.g., salary increases, stock options). Welch (2001, p. 160) states that "A's should be getting raises that are two to three times the size given to the B's." Levinson (2003) states that "stars in the top group receive the lion's share of development and bonuses." Although there is little research on differentiation of rewards in FDS, given that GE and Welch have probably been the most influential proponents of FDS, we relied on his recommendation. Therefore, we decided to make our lower reward differentiation for A's 1–2 times more than B's and our higher reward differentiation for A's 3–4 times more than B's.

Size of Comparison Group

Comparison group size in FDS likely vary based on the industry, company size, and job class. Although to the authors knowledge there is no research on this element within FDS, research on managers' span of control indicates that the average number of direct reports is seven, although certain industries and larger organizations may have an average of nine to sixteen (Davison 2003). This suggests that if an organization would ask each manager to rate his or her direct reports, the most relevant comparison group size may be around ten, which is the value of our lower comparison group size.

In organizations such as GE where several managers meet to discuss and rank employees in a similar job class, the comparison group size would likely be larger. Axelrod et al. (2002) suggest that groups should have at least 30 people so they reflect the typical range of performance levels in the company. Bates (2003) also gives an example of assessors from a large consumer packaged-goods company that forced-ranked 37 individuals, resulting in the assignment of 7 A's, 26 B's, and 4 C's. Grote (2005) gives

an example where 47 employees were being reviewed. Based on these examples and other evidence, we believe that a realistic higher comparison group size would be around 50.

Frequency and Consistency of Feedback

Typically, managers are required to give formal feedback once per year, often in conjunction with compensation adjustments (Zetlin 1994). Many experts recommend providing frequent feedback, such as biannually or quarterly (London 2003). A recent survey that includes responses from Canada's 1,000 largest companies found that 50, 27 and 14% of managers conducted formal performance appraisals annually, biannually and quarterly, respectively (Milne 2002). Therefore, we used the most typical annual feedback as the lower condition and 3–4 times per year as the higher condition for frequency of feedback.

Measures

Cognitive Ability

Cognitive ability was measured with the Wonderlic Personnel Test (Wonderlic Personnel Test Manual 1983). This 12-min, standardized intelligence test was completed by all participants prior to all other measures, at an earlier point in the semester. It is correlated (range = .85–.93) with the Wechsler Adult Intelligence Scale full scale (Dodrill 1981; Dodrill and Warner 1988) and has shown strong test–retest reliability (Dodrill 1983) and validity (McKelvie 1989). Normative data indicate that the mean score for the first-year college students is 24 out of 50; the mean for this sample was 27.

Attraction to FDS

Attraction to FDS was measured by asking respondents how attractive it would be to work for the company described in each of the scenarios (i.e. I would be attracted to work for this company) on a 5-point Likert scale ranging from 'strongly agree' to 'strongly disagree'. In order to add increased fidelity to the situation, we asked respondents to rate their attraction to the company rather than the FDS itself. Important to note, however, is that the FDS information was the only information about the company available to respondents. Thus, it is reasonable to assume that the only reason respondents would indicate a given level of attraction to the company is due to the information presented regarding the FDS.

This measure is general in nature, similar to the item "how likely is it that you would pursue interviewing with this organization" that Cable and Judge (1994) used in

their policy capturing study examining respondents' attraction to organizations based on pay system characteristics. Also, similar to Aiman-Smith et al. (2001), in this study attraction can be considered to be an attitude or expressed affect toward a FDS.

Analytic Strategy

In a policy-capturing design, there are data at the within-subject level of analysis (where each subject's decision policy is captured) and at the between-subject level of analysis (where the focus is on the impact of decision-maker characteristics on decision policies). We analyzed the data using hierarchical linear modeling (HLM; Bryk and Raudenbush 1992). The technique has recently been advocated for policy-capturing data because it allows a parsimonious examination of within- and between-person variance (Mellor et al. 1999; Morrison and Vancouver 2000). In the Level 1 (within-subject) analysis, ordinary least squares (OLS) regression equations were calculated for each individual by regressing attraction to FDS on the four FDS elements. This allowed us to pool the element coefficients (beta weights) to determine the average importance of each cue across individuals.

The Level 2 (between-subject) analysis used a restricted maximum likelihood approach in which the intercept and slope coefficients estimated in the Level 1 model were regressed onto Level 2 predictors (i.e. gender, cognitive ability and business major). This analysis enabled us to determine if between person variance in the intercept and slope coefficients could be predicted by individual difference variables. In other words, did the individual difference variables moderate the relationship between the FDS elements and respondents' attraction to FDS?

Results

Descriptive statistics and correlations for all measures are reported in Table 2. The mean rating across all scenarios was 3.20, indicating moderate attraction overall to FDS for this sample. The significant correlations in Table 2 indicate that business majors have higher cognitive ability than non-business majors, while males were more likely than females to be business majors in our sample. To examine respondents' inter-rater reliability between the scenarios, four random scenarios were replicated (Aiman-Smith et al. 2002). Reliability between the scenarios was assessed by examining the relationship between the responses of each of the four duplicated scenarios. The average reliability coefficient was .72. In addition, a *t*-test between the duplicate responses indicated no significant differences. These analyses indicated good reliability and that the

Table 2 Correlations between attraction to FDS and individual differences

Variable	<i>M</i>	<i>SD</i>	1	2	3	4
Level 1						
Attraction to FDS ^a	3.20	.56				
Level 2 ^b						
Gender ^c	1.33	.47	-.03	–		
Cognitive ability ^d	27.2	4.75	.00	-.14	–	
Business major ^e	.71	.46	.03	-.24**	.21**	–

^a Average response across all 16 scenarios

^b *N* = 163 subjects, *N* = 2,608 observations

^c 1 = male, 2 = female

^d Measured using the Wonderlic Personnel Test

^e 0 = Non-business major, 1 = Business major

** *p* < .01

respondents generally took the task seriously and responded consistently to the scenarios.

Hypotheses 1–4 predicted that individuals' attraction to FDS would be related to the four FDS elements. To determine the amount of variance explained by the four FDS elements, we first ran a null model—where the outcome variable was regressed on a unit vector and no parameters are selected (Hofmann 1997). We then computed the *R*² value for the Level 1 predictors as the total variance (i.e., estimated by the null model) minus the variance not attributable to the Level 1 predictors, divided by the total variance. The calculated effect size measure (see Table 3) indicates that the set of FDS elements averaged across subjects accounted for 54.1% of the explainable within-subjects Level 1 variance in the dependent variable by the four FDS elements (Bryk and Raudenbush 1992). The estimates of the average intercept and slopes across individuals are also reported in Table 3. The average slope coefficients, or regression weights, for each of the FDS elements differed significantly from zero.

Each of the four elements of FDS was used by respondents in making decisions about how attracted they were to FDS.² Based on the standardized weights in Table 3, we were able to calculate the average relative importance of the FDS elements on respondents' decision policies. Participants paid most attention to the consequences of poor performance (46%). This element was about twice as important as the next important element, feedback frequency (25%). Participants placed the least emphasis on comparison group size (17%) and reward differentiation (12%).

² We did not hypothesize interactions, and post hoc analyses revealed no statistically significant interactions were present.

Table 3 Level 1 model of FDS elements on attraction to FDS

Variable	Attraction to FDS			
	β	SE ^a	<i>t</i>	Variance ^b
Intercept	3.20**	.04	72.99	.29**
Consequences of poor performance (i.e. termination)	−.64**	.06	−10.96	.45**
Reward differentiation	.17**	.05	3.30	.32**
Comparison group size	.23**	.05	4.31	.35**
Feedback frequency	.34**	.04	8.52	.15**
Effect size (%) ^c				54.1

N = 163

^a Average estimated SE of the Level 1 regression coefficients

^b Variance in Level 1 parameter estimates and chi-square test of significance of variance

^c Percentage of explainable within-subjects Level 1 variance in the dependent variable accounted for by the four FDS elements

** *p* < .01

Hypothesis 1 stated that more stringent consequences for poor performers would be negatively related to how attracted individuals are to FDS. The negative direction of the beta coefficient for consequences of poor performance supports this hypothesis. Support was found for hypothesis 1 such that participants’ attraction to a FDS decreased when the consequences for poor performance were higher (or more stringent). Hypothesis 2 predicted that higher levels of reward differentiation would be negatively related to how attracted individuals are to FDS. Results showed that higher levels of reward differentiation were positively related to how attracted individuals were to FDS, and thus hypothesis 2 was not supported. Hypotheses 3 and 4 stated that larger comparison groups and more frequent feedback, respectively, would be positively related to how attracted individuals are to FDS. The positive direction of the beta coefficients support these hypotheses and indicate that as comparison group size and feedback frequency increase, participants’ were more attracted to FDS. For example, as the frequency of feedback increases from one time per year to 3–4 times per year, respondents were more attracted to FDS.

Table 4 provides the means and standard deviations for the four elements by two levels across all participants. The mean rating (i.e., 3.20) increased or decreased depending on the condition. For example, holding the other three elements constant, the ‘Consequences of Poor Performance’ element either increased or decreased the average level of attraction to FDS to 3.52 (i.e., 3.2 + .32) in the lower condition versus 2.88 (i.e., 3.2 − .32) in the higher condition.

We can also examine the average attractiveness of the two most extreme scenarios. The scenario where there are higher (or more stringent) consequences for poor performers, lower reward differentiation, lower comparison group size, and lower feedback frequency had an average

Table 4 Means and standard deviations^a of attraction to FDS for each FDS element by condition

Variable	Attraction to FDS	
	Mean rating for lower condition	Mean rating for higher condition
Consequences of poor performance (i.e. Termination)	3.52 (.23)	2.88 (.22)
Reward differentiation	3.12 (.37)	3.29 (.38)
Comparison group size	3.08 (.37)	3.32 (.37)
Feedback frequency	3.03 (.35)	3.37 (.35)

N = 163

^a Standard deviations in parentheses

rating of 2.51 (i.e., 3.2 − .32 − .085 − .115 − .17). On a 5-point scale, this rating indicates that respondents slightly disagreed that they were attracted to this description of FDS. On the other hand, the scenario where there are lower (or less stringent) consequences for poor performers, higher reward differentiation, higher comparison group size, and higher feedback frequency had an average rating of 3.89 (i.e., 3.2 + .32 + .085 + .115 + .17). On a 5-point scale, this rating indicates that, on average, respondents agreed that they were attracted to this description of FDS.

Before we proceeded to the Level 2 analyses, we needed to determine if there was systematic variance across the Level 1 slopes and intercepts. Table 3 presents the significant random effects for the slopes and intercepts at Level 1. The results illustrate that there is significant systematic variance in the intercepts and slopes across individuals. We also calculated a residual intra-class correlation of 22%, which reflects the portion of total variance remaining that can be explained by individual differences.

We proceeded to model this variance by using Level 2 predictors to explore whether these individual differences would influence the emphasis that participants placed on each of the four elements of FDS. We regressed the Level 1 slope coefficients for the FDS elements onto gender, cognitive ability, and business major. Table 5 contains these results as well as a measure of effect size. Although the effects of the individual difference variables appear to be relatively small, the reported percentages are not directly comparable with the R^2 statistic. Instead, they measure the fraction of explainable variation remaining to be explained by individual differences (Kristof-Brown et al. 2002). In this case, a total of 20% (out of the possible 22%) of the variance was explained by the individual difference variables, suggesting that the significant results are robust.

Table 5 Results of hierarchical linear modeling level 2 analyses for individual differences

Variable	Attraction to forced distribution		
	β Coefficient	SE	t
Consequences of poor performance (i.e. termination)			
Intercept	-.64**	.06	-11.13
Business major	.03	.13	.23
Gender	-.34**	.13	-2.70
Cognitive ability	.00	.01	.41
Effect size (%) ^a			1
Reward differentiation			
Intercept	.17**	.05	3.40
Business major	-.18	.11	-1.56
Gender	-.16	.11	-1.45
Cognitive ability	.02*	.01	2.06
Effect size (%) ^a			12
Comparison group size			
Intercept	.23**	.05	4.28
Business major	.05	.12	.44
Gender	.04	.12	.38
Cognitive ability	.01	.01	1.12
Effect size (%) ^a			0
Feedback frequency			
Intercept	.34**	.04	8.56
Business major	-.16	.09	-1.74
Gender	.12	.09	1.36
Cognitive ability	.01	.01	.68
Effect size (%) ^a			7

$N = 163$

^a Percentage of explainable Level 2 variance in the dependent variable accounted for by business major, gender and cognitive ability

** $p < .01$

* $p < .05$

The relationship between the consequences of poor performance and attraction to FDS was moderated by gender ($\beta = -.34$; $p < .01$). Females were less attracted than males to FDS when the consequences of poor performance were high (e.g., termination). In addition, the relationship between reward differentiation and an individual's attraction to FDS is stronger for respondents with higher cognitive ability ($\beta = .02$; $p < .05$). In other words, respondents with higher cognitive ability placed more emphasis on high reward differentiation than did respondents with lower cognitive ability. For feedback frequency and comparison group size, neither gender nor cognitive ability was significantly associated with variance in the slope for the outcome variable. This indicates that both respondents with lower and higher cognitive ability and both males and females viewed these elements similarly. In addition, business major was not significantly associated with variance in the slope of attraction to FDS for any of the four elements of FDS, suggesting that both business and non-business majors viewed each of the elements similarly.

Discussion

What stands out most in these findings is the significant effect of different elements of FDS design on the perceived attraction to such a performance management system in place. That is, when college students about to enter the workforce were presented with a full array of different manifestations of FDS, clear and systematic preferences emerged. More specifically, respondents were most inclined to find attractive those systems that had less stringent consequences for low performers, higher differentiation of rewards, large comparison groups, and frequent feedback. The consequences for low performers had the single most powerful influence on their attraction to different FDS. In addition, those with high cognitive ability particularly favored high reward differentiation and males were considerably less affected by more stringent consequences for low performers than females.

Clearly, all FDS are not perceived the same and the elements of attraction and aversion are both interesting and practically important. In the present study, all four manipulated elements of FDS significantly influenced subjects' attractiveness perceptions of a FDS. Below we elaborate on the findings related to each hypothesis as well as our subsequent analysis of the effects of gender and cognitive ability.

Support for hypothesis 1 was strong and indicates that the consequence for lower performers was the most important decision criterion for this sample. This suggests that how low-rated employees are treated may well be the

most sensitive and potentially “culture killing” variable associated with FDS. Even within a sample drawn from a top-ranked business school, presumably laden with talented and high-potential candidates, there was still a notable aversion to negative consequences for low performers.

Hypothesis 2 proposed that ratees would be less attracted to greater levels of differentiation in rewards but, in fact, the findings supported the *opposite* direction in that they were more attracted to greater levels of differentiation in rewards. The hypothesis was based on accounts from the field where wide difference in pay, particularly among people with the same or similar job requirements, can be perceived as inequitable and have negative effects on performance (Bloom 1999). However, for this sample, higher reward differentiation was significantly more attractive.

Although the reward differentiation element had the smallest impact on participant ratings, participants were more attracted to systems with substantive differences in reward allocations as opposed to a system that does not have as large of differentiation between rewards. This inverted finding for hypothesis 2 is perhaps not as surprising given the sample. These were students primed to graduate from an American business school and made up largely of US-born citizens. Western-oriented models of motivation would predict that perceptions of lower payoffs for top performance would result in relatively lower motivation and performance, particularly among those who perceive that they have the ability to perform at high levels. Whether such findings would be replicated with participants from other cultures, or employees with more experience in organizational contexts, is an important conceptual and empirical question.

Hypothesis 3 and hypothesis 4 proposed that larger comparison group size and more feedback would have a significant influence on respondents’ choice of the most attractive FDS profiles. In both cases those hypotheses were supported and thus are relevant variables in how this population views FDS attractiveness. However, it is also notable that both variables were considerably less influential in choices made than were consequences for low performers.

Although it may seem self-evident that larger comparison groups are better, a key point of demarcation is whether the comparison group should extend beyond the span of control of individual managers. While such enlargement will enhance the probability of a normal distribution, it also inevitably creates an evaluation context whereby raters are inherently less familiar with ratee performance. Given that most managers have only 5–15 direct reports (Davison 2003), a pragmatic question becomes how far beyond the average manager’s span of control it is reasonable to go to attain a large comparison group.

Similarly, although it is hardly provocative to find that people prefer contexts where they receive more feedback, it is important to recognize that feedback commands a commitment of management time and expertise. Further, as the size of the comparison group increases beyond the span of control of individual managers, it becomes increasingly difficult and complex to create a context where all people in a comparison group are given frequent, consistent, and meaningful feedback. So, while these data confirm the intuitive notion that feedback amount and frequency are important, further specification of the *relative* importance is of practical concern. That is, if feedback is not of relatively high importance to the perceived attractiveness of FDS, then other issues may rightly assume more attention in the design and execution of such systems.

Dividing the analysis by gender revealed that females were particularly averse to FDS when the consequences of poor performance were more stringent. This is consistent with recent research that has found that males may be more likely than females to embrace a competitive environment where there are “winners” and “losers.” A study by Robinson and Lipman-Blumen (2003), who collected data from 1984 to 2002 on 2,371 males and 1,768 female US managers, supports the notion that males prefer competitive situations more than females do. They measured nine leadership styles by asking how frequently individuals call upon certain behaviors to reach their goals. They found that although men and women exhibited similar leadership styles in six of the nine achieving styles, the most pronounced difference was that male managers exhibited more competitive behaviors than female managers. Therefore, organizations that are attempting to attract a diverse applicant pool may be wise to examine how they deal with low-rated employees as one signal of their culture.

Moreover, based on the analysis of individual differences, respondents with higher cognitive ability placed more emphasis on high reward differentiation than did respondents with lower cognitive ability. This finding is generally consistent with both prior research and managers’ intuitive notions. That is, the data support an increased importance of meaningful reward distinctions for those highest in cognitive aptitude. Because high achievers have strong tendencies toward competitiveness and comparative performance (Kanfer and Heggestad 1997), it is not surprising that they have stronger preferences for high differentiation of rewards. Therefore, reward differentiation may be especially important to attracting employees with high potential and building the talent level in the organization (Trank et al. 2002).

With respect to future discussions of FDS, the present findings confirm that: (1) all FDS are not alike and (2) these are at least four salient design elements that matter. Our hope is that future accounts (academic and popular press)

will more carefully clarify the nature of these particular elements when reporting FDS. Such precision will be useful in generating an increased knowledge base in this area.

For those interested in effectively implementing FDS, the findings suggest that it would be prudent to consider all four elements explored here—with the treatment of low performers being the most salient issue. In fact, the current findings may be useful in helping to derive something of an effective middle ground that enables achievement of some of the noble advantages of FDS without some of the attendant problems. That is, the present data suggests that a FDS that was implemented with developmental options for low performers, an emphasis on consistent feedback and only in those contexts with sufficient comparison group size, would have the highest likelihood of successful application, at least among a young entering workforce.

Limitations and Future Research

In interpreting these findings at least three limitations of the study warrant specific mention. First, results from a sample of college students with limited work experience may not generalize to other samples of the workforce that have more work experience. Also, while college students nearing graduation are an important recruiting target for organizations and thus important for understanding issues of attraction, they are not appropriate subjects for understanding job satisfaction, motivation, or perceptions which require significant job experience. Thus, no inferences can be drawn regarding retention, performance or other variables potentially influenced by FDS elements.

Second, the present study is limited to perceptual data and we want to be cautious not to overstate the practical importance. While this study has moved us forward in our understanding of what young, high potential candidates find attractive in FDS, an important next step is to include the collection of performance outcome data related to performance evaluation systems.

Third, we recognize that the design, implementation, and evaluation of a FDS are more complicated than this study implies. For example, FDS may be used in conjunction with performance evaluation information or it may be combined with other methods, such as supervisors presenting detailed performance ratings and descriptions about their subordinates for discussion and ranking by the supervisor team before the forced distribution is determined. Who provides the rating (e.g., immediate supervisor with or without the input of his supervisor, the peer group, etc.) may also influence perceptions for the FDS. Finally, competition for talent in the labor market and similarity of jobs could also influence perceptions of FDS. In sum, there are a number of variables that could influence perceptions

of FDS, and we only claim to have identified a core set of these elements of FDS.

The limitations of the present study highlight the reality that the future research needs are varied and great, and three specific areas seem most in need of research attention. First, this study highlights the real need to get beyond simple categorical descriptions and prescriptions regarding FDS. Two important considerations are (a) which job levels of the organization might FDS be most applicable (e.g., broad vs. narrow focus) (Dreher and Dougherty 2002)? and (b) for what purpose or for how long (e.g., short-term vs. long-term) should FDS be implemented (Scullen et al. 2005)? For example, FDS may be used for a variety of personnel decisions including retention, promotion, identification of talent for placement in a fast track development program, and/or compensation. Another difference pertains to whether the distribution guidelines for each rating category are strictly enforced or simply given as a recommendation. Finally, some have suggested that a critical element in the effectiveness of such systems is likely to be the existing organizational culture (Guralnik et al. 2004). Clearly, in addition to differences in the design of FDS, a more complex and nuanced understanding of the different ways and contexts in which FDS are actually implemented would be an important contribution.

Second, there is an important need to increase our database of outcomes of different performance management systems in typical organizational contexts and among diverse workforce populations. As organizations employ an increasingly global workforce, there is a pressing need for evidence regarding the effects of HR systems in general, and FDS in particular, in different cultures and settings. For example, how FDS are perceived by young Chinese or Indian or Brazilian graduates is a simple, but interesting extension of the present work. It would also be interesting to examine how preferences for absolute versus comparative systems such as FDS may differ in diverse contexts. More generally, a greater understanding of how FDS impact cross-cultural metrics related to attraction, motivation, retention and performance is an important research pursuit.

Finally, there remains a conspicuous gap in the empirical data with respect to the perceptions and behaviors of the *raters* (not just ratees) involved in FDS (as an exception, see Schleicher et al. 2009). Gaining an understanding of how those who are responsible for actually conducting the ratings, providing the feedback, and managing group morale and performance are essential for the future of FDS research. For example, determining the extent to which raters, ratees and senior managers agree (or disagree) with respect to the perceived importance of different elements is one specific and important research direction that stems from the present study.

Conclusion

While there is much to be done, the exciting news is that there are few areas today where research holds the potential for such immediate practical impact. Managers in every organization are charged with delivering higher economic returns in an increasingly competitive business environment and are challenged with how best to attract and motivate high performing people. Many high-profile and progressive firms continue to espouse the importance, in whole or in part, of FDS and it is therefore important to seek the most effective variants of these systems. Firms are naturally drawn to the success of companies like GE that have successfully implemented FDS, and yet are also eager to avoid an unsuccessful introduction of an FDS. Greater levels of rigorous empirical data will be critical to inform that tension.

Acknowledgments Previous versions of this paper were presented at the 2006 Academy of Management Conference in Atlanta, GA and the 2007 Society for Industrial and Organizational Psychology in Philadelphia, PA. We gratefully acknowledge the helpful comments of George Dreher.

References

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental psychology* (Vol. 2, pp. 267–299). New York: Academic Press.
- Aiman-Smith, L., Bauer, T. N., & Cable, D. M. (2001). Are you attracted? Do you intend to pursue? A recruiting policy capturing study. *Journal of Business and Psychology*, *16*, 219–237. doi:10.1023/A:1011157116322.
- Aiman-Smith, L., Scullen, S. E., & Barr, S. H. (2002). Conducting studies of decision making in organizational contexts: A tutorial for policy-capturing and other regression-based techniques. *Organizational Research Methods*, *5*, 388–414. doi:10.1177/109442802237117.
- Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology*, *49*, 141–168. doi:10.1146/annurev.psych.49.1.141.
- Axelrod, B., Handfield-Jones, H., & Michaels, E. (2002). A new game plan for C players. *Harvard Business Review*, *83*, 80–88.
- Bates, S. (2003). Forced ranking. *HRMagazine*, *48*, 62–68. (June).
- Bloom, M. (1999). The performance effects of pay dispersion on individuals and organizations. *Academy of Management Journal*, *42*, 25–40. doi:10.2307/256872.
- Blume, B. D., Baldwin, T. T., & Rubin, R. S. (2005). *Forced ranking: Who is attracted to it? A study of performance management system preferences*. Paper presented at the annual meeting of the Academy of Management, Honolulu.
- Bossidy, L., & Charan, R. (2002). *Execution: The discipline of getting things done*. New York: Random House.
- Breaugh, J. A., & Starke, M. (2000). Research on employee recruitment: So many studies, so many remaining questions. *Journal of Management*, *26*, 405–434. doi:10.1177/014920630002600303.
- Bretz, R. D., & Judge, T. A. (1994). The role of human resource systems in the job applicant decision processes. *Journal of Management*, *20*, 531–551. doi:10.1016/0149-2063(94)90001-9.
- Bretz, R. D., Jr., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management*, *18*, 321–352.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Cable, D. M., & Judge, T. A. (1994). Pay preferences and job search decisions: A person-organization fit perspective. *Personnel Psychology*, *47*, 317–348. doi:10.1111/j.1744-6570.1994.tb01727.x.
- Cable, D. M., & Judge, T. A. (1996). Person-organization fit, job choice decisions, and organizational entry. *Organizational Behavior and Human Decision Processes*, *67*, 294–311. doi:10.1006/obhd.1996.0081.
- 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. *The Journal of Applied Psychology*, *83*, 615–633. doi:10.1037/0021-9010.83.4.615.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *The Journal of Applied Psychology*, *86*, 425–445. doi:10.1037/0021-9010.86.3.425.
- Colvin, G. (2001). We can't all be above average. *Fortune*, *144*, 3.
- Davison, B. (2003). Management span of control: How wide is too wide? *The Journal of Business Strategy*, *24*, 22–29. doi:10.1108/02756660310494854.
- Dodrill, C. B. (1981). An economical method for the evaluation of general intelligence in adults. *Journal of Consulting and Clinical Psychology*, *4*, 668–673. doi:10.1037/0022-006X.49.5.668.
- Dodrill, C. B. (1983). Long term reliability of the Wonderlic Personnel Test. *Journal of Consulting and Clinical Psychology*, *51*, 316–317. doi:10.1037/0022-006X.51.2.316.
- Dodrill, C. B., & Warner, M. H. (1988). Further studies of the Wonderlic Personnel Test as a brief measure of intelligence. *Journal of Consulting and Clinical Psychology*, *56*, 145–147. doi:10.1037/0022-006X.56.1.145.
- Dreher, G., & Dougherty, T. W. (2002). *Human resource strategy: A behavioral perspective for the general manager*. Irwin/McGraw-Hill.
- Duffy, K. E., & Webber, R. E. (1974). On “relative” rating systems. *Personnel Psychology*, *27*, 307–311. doi:10.1111/j.1744-6570.1974.tb01536.x.
- Folger, R., & Konovsky, M. A. (1989). Effects of procedural and distributive justice on reactions to pay raise decisions. *Academy of Management Journal*, *32*, 115–130. doi:10.2307/256422.
- Gary, L. (2001). The controversial practice of forced ranking. *Harvard Management Update*, *October*, 1–2.
- Gerhart, B., & Milkovich, G. T. (1990). Organizational differences in managerial compensation and financial performance. *Academy of Management Journal*, *33*, 663–691. doi:10.2307/256286.
- Gerhart, B., & Milkovich, G. T. (1992). Employee compensation: Research and practice. In M. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 3, pp. 481–569). Palo Alto, CA: Consulting Psychologists Press.
- Gladwell, M. (2002). The talent myth. Are smart people overrated? *New Yorker (New York, N.Y.)*, (July), 22.
- Goffin, R. D., Gellatly, I. R., Paunonen, S. V., Jackson, D. N., & Meyer, J. P. (1996). Criterion validation of two approaches to performance appraisal: The behavioral observation scale and the relative percentile method. *Journal of Business and Psychology*, *11*, 23–34. doi:10.1007/BF02278252.
- Grote, D. (2002). Forced ranking: Behind the scenes. *Across the Board*, 40–45 (Nov./Dec).

- Grote, D. (2005). *Forced ranking: Making performance management work*. Boston: Harvard Business School.
- Guralnik, O., Rozmarin, E., & So, A. (2004). Forced distribution: Is it right for you? *Human Resource Development Quarterly*, *15*, 339–345. doi:10.1002/hrdq.1107.
- Hedge, J. W., & Teachout, M. S. (2000). Exploring the concept of acceptability as a criterion for evaluating performance measures. *Group & Organization Management*, *25*, 22–44. doi:10.1177/1059601100251003.
- Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology*, *39*, 811–826. doi:10.1111/j.1744-6570.1986.tb00596.x.
- Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, *23*, 723–744. doi:10.1177/014920639702300602.
- Huselid, M. A., Beatty, R. W., & Becker, B. E. (2005). A players or a positions? *Harvard Business Review*, *83*, 110–117.
- Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, *50*, 905–925. doi:10.1111/j.1744-6570.1997.tb01487.x.
- Judge, T. A., & Bretz, R. D. (1992). Effects of work values on job choice decisions. *The Journal of Applied Psychology*, *77*, 261–271. doi:10.1037/0021-9010.77.3.261.
- Judge, T. A., & Ferris, G. R. (1993). Social context of performance evaluation decisions. *Academy of Management Journal*, *36*, 80–105. doi:10.2307/256513.
- Jurgensen, C. E. (1978). Job preferences (what makes a job good or bad?). *The Journal of Applied Psychology*, *63*, 267–276. doi:10.1037/0021-9010.63.3.267.
- Kanfer, R. (1990). Motivational theory and industrial and organizational psychology. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 75–105). Palo Alto, CA: Consulting Psychologist Press.
- Kanfer, R. M., & Heggstad, E. D. (1997). Motivational traits and skills: A person-centered approach to work motivation. In B. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 19, pp. 1–56). Greenwich, CT: JAI Press.
- Karren, R. J., & Barringer, M. W. (2002). A review of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods*, *5*, 337–361. doi:10.1177/109442802237115.
- Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *The Journal of Applied Psychology*, *85*, 708–723. doi:10.1037/0021-9010.85.5.708.
- Korsgaard, M. A., & Roberson, L. (1995). Procedural justice in performance evaluation: The role of instrumental and non-instrumental voice in performance appraisal discussions. *Journal of Management*, *21*, 657–669. doi:10.1177/014920639502100404.
- Kristof-Brown, A. L., Jansen, K. J., & Colbert, A. E. (2002). A policy capturing study of simultaneous effects of fit with jobs, groups, and organizations. *The Journal of Applied Psychology*, *87*, 985–993. doi:10.1037/0021-9010.87.5.985.
- Kruger, A. N., & DeNisi, A. S. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284. doi:10.1037/0033-2909.119.2.254.
- Landy, F. J., Barnes, J. L., & Murphy, K. R. (1978). Correlates of perceived performance and accuracy of performance evaluation. *The Journal of Applied Psychology*, *63*, 751–754. doi:10.1037/0021-9010.63.6.751.
- Lawler, E. E. (2002). The folly of forced ranking. *Strategy + Business*, *28*, 28–32.
- Lawler, E. E. (2003). Reward practices and performance management effectiveness. *Organizational Dynamics*, *32*, 396–404. doi:10.1016/j.orgdyn.2003.08.007.
- Levinson, M. (2003). One tough job: How to find, fix or fire your poor performers; bad employees drain your IT organization and the company. Forced ranking can help you get tough—but at what cost? *CIO*, *17*(3), 1–4.
- Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*, 881–905. doi:10.1016/j.jm.2004.06.005.
- London, M. (2003). *Job feedback*. Upper Saddle River, NJ: Earlbaum.
- McBriarty, M. A. (1988). Performance appraisal: Some unintended consequences. *Public Personnel Management*, *17*, 421–434.
- McKelvie, S. J. (1989). The Wonderlic Personnel Test: Reliability and validity in an academic setting. *Psychological Reports*, *65*, 161–162.
- Mellor, S., Paley, M. J., & Holzworth, R. J. (1999). Fans' judgments about the 1994–95 major league baseball players' strike. *Multivariate Behavioral Research*, *34*, 59–87. doi:10.1207/s15327906mbr3401_3.
- Milne, J. L. (2002). Checking in: Survey shows frequency of employee performance reviews. *Canadian Manager*, *27*, 2.
- Morrison, E. W., & Vancouver, J. B. (2000). Within-person analysis of information seeking: The effects of perceived costs and benefits. *Journal of Management*, *26*, 119–137. doi:10.1016/S0149-2063(99)00040-9.
- Mount, M. K. (1984). Satisfaction with a performance appraisal system and appraisal discussion. *Journal of Occupational Behavior*, *5*, 271–279. doi:10.1002/job.4030050404.
- Murphy, K. R., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Nathan, B. R., & Alexander, R. A. (1988). A comparison of criteria for test validation: A metal-analytic investigation. *Personnel Psychology*, *41*, 517–535. doi:10.1111/j.1744-6570.1988.tb00642.x.
- Olson, C. A., & Davis, G. M. (2003). Pros and cons of forced ranking and other relative performance ranking systems. *Society for Human Resource Management Legal Report*, March.
- Organ, D. W. (1990). The subtle significance of job satisfaction. *Clinical Laboratory Management Review*, *4*, 94–98.
- Pfeffer, J. (1998). *The human equation: Building profits by putting people first*. Boston: Harvard Business School.
- Pfeffer, J. (2001). Fighting the war for talent is hazardous to your organization's health. *Organizational Dynamics*, *29*, 248–259. doi:10.1016/S0090-2616(01)00031-6.
- Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review*, *84*, 62–74.
- Robinson, J. L., & Lipman-Blumen, J. (2003). Leadership behavior of male and female managers, 1984–2002. *Journal of Education for Business*, *79*, 28–33.
- Roch, S. G., Sternburgh, A. M., & Caputo, P. M. (2007). Absolute vs. relative performance rating formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment*, *15*, 302–316. doi:10.1111/j.1468-2389.2007.00390.x.
- Rousseau, D. M., & Anton, R. J. (1988). Fairness and implied contract obligations in job terminations: A policy capturing study. *Human Performance*, *1*, 273–289. doi:10.1207/s15327043hup0104_4.
- Rynes, S. L., Brown, K. G., & Colbert, A. E. (2002). Seven common misconceptions about human resource practices: Research findings versus practitioner beliefs. *Academy of Management Executive*, *16*, 92–103.

- Schleicher, D. J., Bull, R. A., & Green, S. G. (2009). Rater reactions to forced distribution rating systems. *Journal of Management*, (in press).
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274. doi:10.1037/0033-2909.124.2.262.
- Schwab, D. P., Rynes, S. L., & Aldag, R. J. (1987). Theories and research on job search and choice. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resource management* (Vol. 5, pp. 126–166). Greenwich, CT: JAI Press.
- Scullen, S. E., Bergey, P. K., & Aiman-Smith, L. (2005). Forced distribution rating systems and the improvement of workforce potential: A baseline simulation. *Personnel Psychology*, *58*, 1–32. doi:10.1111/j.1744-6570.2005.00361.x.
- Shirouzu, N. (2001). Ford stops using letter rankings to rate workers. *Wall Street Journal*, B.1 (July 11).
- Taylor, M. S., Masterson, S. S., Renard, M. K., & Tracy, K. B. (1998). Managers reactions to procedurally just performance management systems. *Academy of Management Journal*, *41*, 568–579. doi:10.2307/256943.
- Tichy, N. M., & Sherman, S. (2001). *Control your destiny or someone else will: Lessons in mastering change-from the principles jack welch is using to revolutionize GE*. New York: HarperCollins.
- Trank, C. Q., Rynes, S. L., & Bretz, R. D. (2002). Attracting applicants in the war for talent: differences in work preferences among high achievers. *Journal of Business and Psychology*, *16*, 331–345. doi:10.1023/A:1012887605708.
- Wagner, S. H., & Goffin, R. D. (1997). Differences in accuracy of absolute and comparative performance appraisal methods. *Organizational Behavior and Human Decision Processes*, *70*, 95–103. doi:10.1006/obhd.1997.2698.
- Walster, E., Walster, G. W., & Scott, W. G. (1978). *Equity: Theory and research*. Boston: Allyn & Bacon.
- Welch, J. F. (2001). *Jack: Straight from the gut*. New York: Warner Books, Inc.
- Wonderlic Personnel Test Manual. (1983). Northfield, IL: E.F. Wonderlic & Associates.
- Wright, R. P. (2002). Perceptual dimensions of performance management systems in the eyes of different sample categories. *International Journal of Management*, *19*, 184–193.
- Zetlin, M. (1994). Up for review. *Sales & Marketing Management*, *146*, 82–86.