

Measuring Teacher and Leader Performance

Cross-Sector Lessons for
Excellent Evaluations

PUBLIC IMPACT



JULIE KOWAL AND EMILY AYSCUE HASSEL



About the Authors

JULIE KOWAL is a senior consultant with Public Impact. She consults and leads teams to offer research-based guidance on several challenging policy and management issues in education, including teacher quality, evaluation, retention, and compensation; and dramatic change in persistently underperforming schools. Ms. Kowal recently led Public Impact's involvement in several applications for the federal Race to the Top and Investing in Innovation competitions. She also serves as editor for all Public Impact publications. An alumna of AmeriCorps NCCC and Public Allies DC, Ms. Kowal earned her law degree with honors from the University of North Carolina at Chapel Hill.

EMILY AYSCUE HASSEL is Co-Director of Public Impact. She provides thought leadership and oversight to Public Impact's work on human capital, organizational transformation, and emerging opportunities for dramatic change in pre-K to grade 12 education. Her work has appeared in *Education Week*, *Education Next* and other publications. She previously worked for the Hay Group, a leading human resources consulting firm. Ms. Hassel received her law and master in business administration degrees from the University of North Carolina at Chapel Hill.

About the Series

This report is part of the series *Building an Opportunity Culture for America's Teachers*. To see all the reports in this series, please visit www.opportunityculture.org.

Made possible with the support of:

TheJoyceFoundation

Acknowledgements

This report was made possible by the generous support of The Joyce Foundation. It is part of a series of reports about "Building an Opportunity Culture for America's Teachers." The authors are grateful to Robin Chait of the Center for American Progress, Tim Daly of The New Teacher Project, and Sabrina Laine of the National Comprehensive Center for Teacher Quality for their helpful feedback and insights as the draft took shape. We are also indebted to Daniela Doyle of Public Impact for her significant research assistance and Bryan Hassel for his feedback on early drafts. Dana Brinson oversaw production and dissemination of the report. Finally, we would like to thank Sharon Kebschull Barrett for careful editing, and April Leidig-Higgins for the design of the report.

© 2010 Public Impact, Chapel Hill, NC

Public Impact is a national education policy and management consulting firm based in Chapel Hill, NC. We are a team of researchers, thought leaders, tool-builders, and on-the-ground consultants who help education leaders and policymakers improve student learning in K-12 education. For more on Public Impact and our research, please visit: www.publicimpact.com.

Public Impact encourages the free use, reproduction, and distribution of this working paper for noncommercial use. We require attribution for all use. For more information and instructions on the commercial use of our materials, please contact us at www.publicimpact.com.

Measuring Teacher and Leader Performance

Cross-Sector Lessons for Excellent Evaluations

By Julie Kowal and Emily Ayscue Hassel

EXECUTIVE SUMMARY

For too long, performance measurement systems in education have failed to document and recognize real differences among educators. But a recent national push to use performance evaluations for critical personnel decisions has highlighted the shortcomings of our current systems and increased the urgency to dramatically improve them. As state and local education leaders reform teacher and principal evaluation systems, they can draw from decades of performance measurement research and experience in other sectors to develop more accurate, reliable, and meaningful information about educators' performance.

In this report, we summarize six steps that research and experience from across sectors — including government agencies, nonprofit organizations, and for-profit companies — show are critical for designing an outstanding performance measurement system:

1. **Determine the purposes of performance measurement**, such as informing professional development, promotions, compensation, retention, and dismissals. Engaging top leadership in conversation about these purposes helps ensure that performance measurement systems provide the type and quality of information necessary to guide each decision.

2. **Choose job objectives that align with the organization's mission** to ensure that performance measures and the measurement process capture the critical outcomes and behaviors needed from each employee to achieve the school's, district's, or education provider's mission.
3. **Design performance measures**, including what individuals in each role are expected to contribute and the ways in which they are expected to achieve results. By choosing the right measures, organizations clarify and stimulate staff actions that contribute to success.
4. **Set performance standards** to use as a yardstick for assessing employees' performance, so that both leaders and staff know what good and great performance looks like.
5. **Design the performance measurement process** by determining who will organize and have input into evaluations, using what process, and how often.
6. **Use measurement results to take action**, including making decisions about professional development; promotions and reach extension; career planning; compensation; retention and dismissals; and future recruiting and hiring.



For each of these steps, we share critical research findings and examples from other sectors to guide the design and implementation of performance measurement systems in education. Many of these lessons echo elements that reformers have recently called for in education,¹ such as:

- ▶ **Measuring both the “what” and “how” of educators’ performance.** Cross-sector research reveals clear advantages to including in employees’ evaluations measures of ultimate results as well as the activities that contribute to those results. Combining both student results *and* educators’ skills and competencies will increase the developmental value of evaluations and lay the groundwork for better-informed decisions about compensation, promotion, and dismissal.
- ▶ **Conquering fears of measuring student learning.** District and state leaders must also consider data about teachers’ and leaders’ contributions to student progress. After all, schools are in the business of learning — and students’ academic growth is the primary outcome. Rather than fear student learning results, we must measure them — and commit to improve our measurement over time rather than using imperfections as an excuse to resist such data entirely.
- ▶ **Using performance measurement results.** Improving our performance measurement systems will have no effect unless the results are put to use. States, districts, and other providers must commit to act upon real differences in educator effectiveness — by using evaluations as the basis for teachers’ and principals’ ongoing development and pay, decisions about retention and dismissal, and future recruiting and selection.

These are commonplace, well-traveled approaches in other sectors. Indeed, there are few compelling reasons or existing barriers to prevent us from implementing them immediately in education. In the final section of this report, we put these lessons to use by providing sample elements of a strong performance

measurement system for teachers, and a detailed example of a method for measuring the “how” of their performance.

Research and experience from other sectors also suggest several ways in which successful organizations have already catapulted past even the boldest calls for reform in education to refine their measurement systems and obtain the most impact, such as by:

- ▶ **Adopting strategies to protect against leniency and bias.** Tactics include rating on a forced distribution, holding managers accountable for the appraisals they conduct, including reviews from multiple perspectives with different weights, and providing training in the performance measurement process.

- ▶ **Meaningfully assessing team performance.** Where individuals’ contributions occur at the team level, great organizations get smart about assessing performance in that context. Methods include identifying individual contributions to team results using peer evaluation and statistical analysis, and assessing team outcomes alongside individual results.

- ▶ **Measuring more to measure better.** Districts, schools, and other providers that are committed to improving the performance of professionals will measure fearlessly and frequently. With a simultaneous commitment to use the lessons and

measurement tools available today and improve them in the future, we can rapidly advance performance measurement — and more quickly contribute to improved student outcomes.

- ▶ **Rapidly adjusting measures that do not predict success.** Great organizations change their measures, and measurement processes, to continue making progress and to reflect changes in how work is done. Education must do no less.

Leaders of successful organizations in other industries have long recognized that talent — and strategic management of that talent — is the key to their competitiveness and success. There is no sector where talent is more important than education. We must carefully consider lessons learned from decades of performance measurement in other sectors, both to inform our current efforts and to spur the next generation of reform.

The steps and decisions involved in performance measurement form the culture in our schools, which in turn determines who decides to teach, how long they stay, what students they reach, and how motivated they feel to raise our children’s sights and conquer obstacles. Measuring educators’ performance accurately — and in ways that enable crucial decisions — must therefore be a top priority for reform within the next half-decade.

Introduction

Our failure to meaningfully differentiate teachers' and principals' impact on student learning is finally beginning to receive the attention it deserves.² For too long, performance measurement systems in education have not documented real differences among educators — in their roles, strengths, weaknesses, or overall effectiveness in improving student learning. However, a recent national push to use performance evaluations for critical personnel decisions has highlighted the shortcomings of our current systems and increased the urgency to dramatically improve them.

How can we make important decisions about teacher and principal assignments, rewards, retention, or dismissal when we cannot reliably recognize differences in effectiveness and impact? How can we offer effective, personalized professional development without accurate and credible information about educators' individual strengths and weaknesses?

Faced with these types of questions, many federal,

state, and district education leaders have begun to realize that our current performance measurement systems not only contribute to a depersonalization of careers in education, but also limit our ability to improve student learning.

We therefore have a great opportunity. As education shifts toward a more performance-focused culture, improving the measurement of teacher and principal effectiveness will be a critical step. With accurate, reliable, and meaningful information about educators' performance, states, districts, and other providers will be able to better differentiate between “effective” and “ineffective” teachers. Accurate and complete measurement will enable better decisions about future recruiting, professional development, career paths, pay, retention, and dismissal.

With better information about educators' performance, the education sector can become a profession in which excellence is recognized and used to help more students achieve success.

As education shifts toward a more performance-focused culture, improving the measurement of teacher and principal effectiveness will be a critical step.



Guideposts from Other Sectors

Lessons about Effective Performance Measurement

Fortunately, decades of performance measurement research and experience in sectors outside education — including government agencies, nonprofit organizations, and for-profit companies — provide strong guideposts for our work in education. Together with a growing knowledge base in the school setting, these cross-sector lessons can help motivated education leaders make better, faster improvements in performance measurement for educators.

In this research summary, we distill the lessons into six elements that cross-sector research and experience show are critical for designing an outstanding performance measurement system. While the report is not a how-to toolkit, we organize these elements into a series of **six steps for effective performance measurement**:

- 1. Determine the purposes of performance measurement**, such as informing professional development; promotions and reach extension; career planning; compensation; retention and dismissals; and future recruiting and hiring. The purposes of a performance appraisal system should determine the rest of its design, including performance goals, measures, and the appraisal process.
- 2. Choose objectives that align with the organization's mission** to ensure that performance measures and the measurement process capture the critical outcomes and behaviors needed from each employee to achieve the school's, district's, or provider's mission.

- 3. Design performance measures**, including what individuals in each role are expected to contribute and the ways in which they are expected to achieve results. By choosing the right measures, organizations stimulate staff actions that lead to success.
- 4. Set performance standards** against which employees' performance will be evaluated by defining what a fully satisfactory performance looks like. This step also includes decisions about whether to use absolute or relative standards, how to account for changing circumstances, and the scales on which each objective will be rated.
- 5. Adopt a performance measurement process** by determining who will organize and have input into evaluations, using what process, and how often.
- 6. Use measurement results to take action**, including making decisions about professional development; promotions and reach extension; career planning; compensation; retention and dismissals; and future recruiting and hiring.

In the final section of the report, we put these lessons to use by providing sample elements of a strong performance measurement system for teachers, and a detailed example of measuring the “how” of their performance. We also recommend several critical considerations for leaders in K–12 education as they overhaul their own measurement systems to foster a strong performance culture in our public schools.

Step One: Identify the purposes of performance appraisals

The first step in developing an effective performance measurement system is to identify the purposes it will serve in the organization, and how the results will be used.³ Most evaluation systems are designed to serve multiple purposes, which require different types of information: for example, decisions about promotion require the ability to make comparisons among employees, while development and training require a deep understanding of individual strengths and weaknesses.⁴ Similarly, appraisals aimed at development may be best conducted separately from those for compensation, because employees may be less likely to embrace feedback if they believe it is tied to their paycheck.⁵ The most common purposes of performance measurement include:

- ▶ **Feedback and development** — to provide information about an employee's performance individually and as a team member, so she is aware of her own strengths and weaknesses, and can focus future work on developing both.⁶ In addition, performance appraisals tell the organization where to focus organization-wide and individualized professional development to improve employees' skills, competencies, and overall performance.⁷
- ▶ **Promotions and career planning** — to identify employees whose performance warrants additional opportunities and responsibilities, formal promotion, or a lateral career move to use and build on strengths; and to keep people focused on additional ways to achieve success and advance their long-term careers.⁸
- ▶ **Compensation** — to recognize, reward and motivate employees through pay awards or salary increases. In the large number of organizations that use differentiated compensation based on performance, data gathered through the performance measurement process are typically the predominant factor in pay decisions.⁹
- ▶ **Retention and dismissals** — to inform fair, reliable, and data-based decisions about which

employees to keep, which require rapid improvement, and which need help seeking opportunities elsewhere.¹⁰

- ▶ **Future recruiting and hiring** — to correlate behaviors, competencies, and skills with performance among current employees, thereby strengthening future decisions about which candidates to recruit and hire.¹¹ In addition, strong performance measurement systems that are linked to rewards have been shown to attract more high-performing and better qualified candidates into an organization's talent pool, creating a virtuous cycle.¹² Good measurement alone is not enough. Prospective candidates must see that it is used to reward better performers.

Engaging top leadership in conversation and deliberation about these varying purposes during the initial design of performance measurement systems helps ensure that the system provides the type and quality of information necessary to inform each decision. An organization's primary goals for the system should direct both the content of the evaluations (discussed in the next section, Step Two) and the process by which appraisals are carried out (covered under Step Four).

Step Two: Choose job objectives that align with the organization's mission

Ultimately, any performance measurement system aims to improve the performance of individual employees and, as a result, improve the overall success of the organization.¹³ To meet this goal, performance measurement systems should be tightly aligned with each employee's expected contributions to the organization's core mission.¹⁴ A 2007 study of historically high-performing companies examined common and disparate characteristics among their performance measurement systems. Among the authors' key findings was that each company had worked to make its system consistent and aligned with its particular culture and mission, to make performance measurement and the organization's goals mutually reinforcing.¹⁵

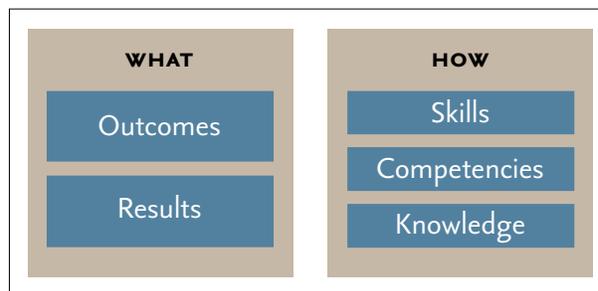
The second critical step successful organizations take in designing a strong performance measurement system, therefore, is to clarify how each job or job category contributes to the organization’s overall mission. This helps ensure that performance objectives capture the critical contributions needed from individuals and teams — i.e., that they are measuring what matters most.¹⁶ Given the goals and direction of the organization, what are employees’ roles in getting there? Without this step, organizations add objectives and process steps that clog the measurement process and miss elements of a job that are essential to the employee’s — and the organization’s — success.

Clarifying how each job contributes to the organization’s mission ensures that performance appraisals measure what matters most.

The types of objectives that organizations often set for individual employees and teams can be grouped into two categories (see Figure 1):¹⁷

- ▶ **The “what”:** What tangible results will the employee be expected to achieve to contribute to the organization’s overall goals? Honing quality, managing costs, increasing sales, and improving speed and innovation — and customer perception of the resulting benefits — are common objectives in the private sector that lead to organization-level success. The relative importance of these objectives depends on the organization’s particular strategy for success. Survey evidence suggests that the objectives that are most easily turned into quantifiable measures predominate in many organizations.¹⁸
- ▶ **The “how”:** What skills, knowledge, and competencies enable the employee to achieve desired results? Objectives of this kind may include achiev-

FIGURE 1. Two Dimensions of Employees’ Performance



ing target levels or improving on a particular skill; improving technical or process knowledge; or demonstrating patterns of thinking and behavior that affect an employee’s ability to achieve results, such as his drive for achievement or ability to influence others.¹⁹

To serve the multiple purposes outlined above, most evaluations focus on both types of objectives.²⁰ Research from across sectors suggests that the best types of both objectives will incorporate several key characteristics. Performance objectives should be:

- ▶ **Designed with employee input.** As suggested under Step One, top leadership in an organization must participate in the design of the performance evaluation system and its purposes. In two-thirds of Fortune 100 companies (the 100 top-ranked U.S. corporations according to gross revenue), decisions about performance appraisal strategy are made by top executives, including whether or not to have appraisals, and the purposes for which they are used.²¹

Just as important, however, is the engagement of employees in the selection of performance objectives and, to the extent possible, how they will be measured against a standard (discussed further under steps Three and Four). Studies of employees across sectors suggest that involving them both in goal setting and feedback can lead to higher levels of acceptance of the appraisal system, higher motivation, and, as a result, increased productivity.²² More than 80 percent of Fortune 100 companies

involve executives, managers, and professional employees in setting their own objectives.²³

- ▶ **Clear and transparent.** Clear and transparent objectives help explain why each element of performance is important and how it fits into the firm's larger goals. They also help communicate to employees exactly what will be expected of them. By visibly aligning organizational purpose and individual performance objectives, organizations can make employees feel more empowered and motivated to deliver expected results.²⁴ In addition, clear performance objectives make the evaluation process easier and more reliable, by signaling to raters how performance should be judged.²⁵
- ▶ **Reflective of differences among positions and roles.** Just as different firms perform different functions, departments within an organization — and employees within those departments — play different roles. Sometimes, employees have standard jobs with performance expectations that are fairly consistent over time and across many people. In other cases, individuals take on roles crafted to fit the firm's current needs and the individual's unique skill combination. In both cases, performance objectives should reflect the actual jobs and roles people have in helping the firm meet its goals. As Dick Grote explains in his seminal *Complete Guide to Performance Appraisal* (1996), “the behaviors, skills and competencies that make an individual a good sales person are likely to be different from those that make for a good clerical employee, just as the attributes of an ideal clerk may differ from those of an ideal supervisor or engineer.” The same can be said for varying jobs and roles within a school.

Indeed, multiple researchers in psychology have found that employees are most likely to be satisfied with ratings that they believe reflect their performance and contributions to the organization. Building this type of rating, however, requires some level of differentiation in how employees within various roles are evaluated.²⁶ Differentiation may be as simple as adjusting the specific

objectives for different positions, or could require evaluating individuals with varying measures in accordance with their role.²⁷

- ▶ **Limited in number.** How many objectives is it reasonable to include for a particular employee's performance appraisal? If the objectives are truly critical to the organization's success, an individual can probably accomplish only a limited number over the course of several months or a year. While there is no hard rule, few successful organizations expect an employee to focus on less than a handful of objectives, or to achieve more than a dozen during a single appraisal period. With too few objectives, the organization is probably not requiring enough from the employee, or is defining objectives too broadly. With too many, it becomes more likely that the employee and the employer will begin to emphasize some and neglect others over time.²⁸ The most important consideration is to concentrate each employee within the organization on the most critical outcomes and actions that will positively affect the success of the firm as a whole.

How many objectives is it reasonable to include in a performance appraisal? If they are truly critical to the organization's success, an individual can probably accomplish only a limited number over the course of several months or a year. More than a few and less than a dozen at one time is a feasible range.

- ▶ **Adjusted to meet changing goals and needs.** Organizations with strong performance measurement systems also design processes that are flexible enough to be adjusted in response to



shifts in the organization's mission or goals, and to changing employee needs. At John Deere, for example, changing needs in the industry in the 1980s required the organization to shift its culture and change its appraisal system as well. This leading agricultural manufacturer had long evaluated and rewarded employees based on individual performance, according to internally developed job standards. But when the agricultural sector plummeted, leaders were pressed to dramatically increase efficiency. A key expectation for manufacturing employees became their ability to work together and use their ingenuity to find more productive work methods. John Deere therefore adjusted its appraisal and reward system to focus on team-based outcomes and successful innovations in accordance with its new goals and strategy.²⁹ The best performance measurement systems are “living” systems, enabling organization leaders to

respond to changing needs, successes, and failures without designing a new system from scratch every few years.

After choosing the objectives that employees are expected to meet, based on the organization's mission and each employee's role in achievement, organizations must next determine how progress on each objective will be measured. This is the focus of Step Three.

Step Three: Design Performance Measures

How do organization leaders know when an employee has met an objective? Sometimes it is obvious, such as when a salesperson turns in dollar figures showing progress toward annual sales goals, or a manager reduces error rates by a certain percentage. But other times there is not a direct measure of an employee's performance relative to a given standard,

such as when the employee is charged with improving communications or encouraging teamwork. In these cases, organizations must choose measures that are precursors and proxies for ultimate performance to provide a best estimate of whether employees have achieved an objective. Below, we examine the processes by which organizations in other sectors make these determinations for objectives in both the “what” and “how” categories.

Measures of “what.” Research suggests that the primary factor in the majority of performance measurement systems is an employee’s tangible and observable results.³⁰ Sales outcomes are the most obvious example of a result measure, because the relationship between the employee’s performance (selling) and the desired outcome (increasing sales) is so direct and the performance so easily measured (sales volume). Where that relationship is direct and straightforward, organization leaders may not think of choosing performance measures as a step distinct from choosing objectives. But other outcomes can be similarly measured and quantified, such as speed (the number of times an employee engages in a particular activity), output (how much product is produced within a given time period), or market share (the individual’s sales revenue as a percentage of all similar purchases by a defined customer group).³¹

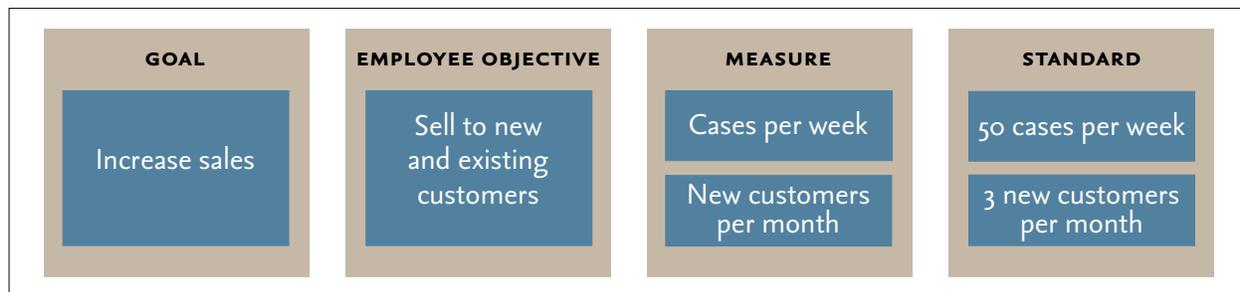
Figure 2 shows the relationship between organizational goals, employee outcomes, and measures of performance for a hypothetical organization, with

related standards for performance (discussed under Step Four).

Measures of “how.” In many cases, an employee is also expected to achieve objectives that precede or are necessary to obtain the ultimate result, such as engaging in teamwork, providing high-quality customer service, or persisting in the face of obstacles. In recent years, organizations across sectors have increasingly begun to use these types of measures as critical elements of employees’ evaluations.³² Particularly in profit-driven industries, organization leaders have recognized that relying upon financial measures alone may undercut a company’s mission or strategy in the longer term, and that measures of *how* employees perform their roles offer a more complete and predictive picture of the organization’s future viability and health.³³ Organizations also have found that purely results-based evaluation models tend to oversimplify what the employee does, and are not helpful for identifying areas for development.³⁴

For example, at banking giant Citibank, financial measures long dominated performance evaluations. In the late 1990s, however, managers realized that these measures failed to connect employees to the bank’s focus on customer service and — more dangerously — tended to focus employees’ energy on short-term targets at the expense of long-term success. To capture the importance of non-financial measures, one state-level division developed an appraisal system that also included measures of the

FIGURE 2. Example of Relationship Between Goals, Objectives, Measures, and Standards



Note: Examples of goals, objectives, measures, and standards for educators appear in the final section of this report, on page 25.

Many organization leaders have recognized that basing performance appraisals upon results-based measures alone can undercut the organization's long-term mission. Including measures of *how* an employee achieves success provides a more accurate picture of performance and helps identify areas for development.

“how,” such as strategy implementation and customer satisfaction. Strategy implementation was measured via revenue and broken down by types of customers that the division especially wanted to serve, such as households, businesses, and professionals. Customer satisfaction was measured through telephone interviews with customers who had visited a branch during the past month. All of these measures could be quantified and compared with performance standards to result in an employee's individual performance rating.³⁵

Rental car agency Enterprise used a similar method to quantify customer satisfaction, which it deemed critical to ensuring repeat business, and, thus, its employees' and organization's success. Branch managers were evaluated in part based on the percentage of customers who rated their branches 5 out of 5 on customer service. To receive compensation bonuses or become eligible for a promotion, the manager's branch was required to deliver customer service at or above the average feedback score for all comparable branches.³⁶

Behavioral competencies are one subset of measures that powerfully examine the “how” of employees' performance. Competencies are habitual behaviors that enable people to achieve results in their

job.³⁷ For example, an employee may achieve outstanding results only by engaging in certain behaviors, habits, or skills such as teamwork, persistence, or flexibility. These are difficult to evaluate objectively unless increasingly effective behaviors are clearly written for managers, peers, customers, and others to score. Measuring an employee's competencies is a critical element of performance measurement, particularly for the purposes of feedback, development, career planning, and future selection.³⁸ Many companies across industries incorporate competencies into performance measurement so that they can more objectively examine the qualitative factors that make employees successful in their jobs.³⁹

For example, Coca-Cola and Mobil Oil both emphasize quantitative (financial and non-financial) results in each employee's performance appraisal. But they also identify the characteristics that they expect from each of their employees in achieving these results, using managers' assessments to evaluate competencies such as communication, problem solving, developing others, technical knowledge, and setting high standards.⁴⁰

Measuring employees' competencies is particularly powerful because it provides a fuller picture of their performance. Competencies correlated with job suc-

FIGURE 3. Competencies Commonly Used in Performance Measurement



cess are remarkably consistent for some positions despite organizational changes and redirection. Many of the same competencies apply across different job categories.⁴¹ Indeed, through decades of rigorous study, researchers have determined that it is possible to identify specific competencies required for success in common job categories, and also to determine statistically the levels needed and relative importance of each competency in distinguishing great from average performers.⁴² A list of competencies that are common performance distinguishers for jobs ranging from managers to sales personnel appears in Figure 3.

Accounting for team-based work. In many organizations, individuals in various roles work together to achieve an important objective. Organizations measure both team effectiveness and individuals' own contributions in this context to make good decisions — both about team assignments and about individual roles. Organization leaders use various strategies to measure team outcomes and individual employees' contributions, including peer reviews, statistical formulas, and team-based results.

When individuals work together to achieve an important objective, organizations measure both team effectiveness and individuals' own contributions to lay the groundwork for good decisions.

► *Peer or 360-degree appraisals* (described in more detail under Step Five, below) enable organization leaders to evaluate each team member's individual contribution to a joint outcome. With feedback from peers and other members of a team about how an employee contributed to the outcome and in what ways, evaluators gain greater insight into the strengths, weaknesses, and contributions of individuals on the team.⁴³ Team members are in a good position to assess one another's competen-

cies, skills, and behaviors because they observe them daily.⁴⁴

At Fortune 500 company Texas Instruments, organization leaders implemented a peer appraisal process to evaluate team-based work. They created a questionnaire describing 33 effective team behaviors with three ratings that team members used to evaluate one another anonymously on a weekly basis. Team members were also asked to rank one another's performance on the team behaviors. The organization then used these rankings to inform decisions about team assignments and bonus pay.⁴⁵

► *Sophisticated statistical formulas* can also help parse out an individual's contribution to a joint outcome. This method is most often used in sports, where the value added by an individual player is determined using statistical formulas. These compare team performance when an individual plays to team performance when that player is not in the game. In addition, since the makeup of players on the team can change, these models also consider how the value of players' contributions changes depending on who is on the field.⁴⁶

► *Team-based results measurement* is also common, and almost always used in combination with peer or individual reviews. Using this strategy, organization leaders hold all employees on a team equally accountable for the team's results. Organizations deliver a uniform performance rating to all members of a team who share responsibility for obtaining results, based on the team's overall achievement of its objectives.⁴⁷

A serious drawback of using team-based results exclusively, however, is what researchers Susanne Scott and Walter Einstein refer to as “social loafing,” which is likely to result in any sector when individual effort is not evaluated and recognized.⁴⁸ Without assessment and action to address loafing, other team members are likely to hold back their efforts as well, leading to a decrease in performance among the team overall. To prevent this phenomenon, appraisals of team performance systems often reflect measures of “what” (such as

productivity, sales volume, or customer satisfaction) and are used in combination with “how” assessments of individual performance.⁴⁹ Many team performance measurement systems, such as those used at technology firm Xerox, are developed jointly by teams, managers, and customers for informal and formal appraisals.⁵⁰

By identifying the key performance measures that underlie successful performance in a job — including the “what” and the “how” that contribute to success, and the manner in which employees work together to achieve it — organizations can define the standards by which employees will be measured. In other words, they can begin to answer the question, “What does excellent performance look like?” In the next sections, we examine the process through which organizations answer the companion question, “How do we know?” by setting standards and designing a performance measurement process.

Step Four: Set Performance Standards

After determining the objectives that employees will be expected to achieve and the measures by which their progress will be evaluated, organization leaders must determine what level of performance on each measure — and overall — is expected by setting performance standards. In education, setting performance standards is controversial, but in most other sectors, it is routine. This step typically includes designing performance categories — the scales to use for rating each measure — and determining which category will serve as the employees’ target, and what is superior or unacceptable performance.

Designing performance categories. Setting performance standards requires first defining the categories into which various levels of performance are grouped. The primary design tasks here are determining the number of categories and choosing their labels, or descriptions that help ensure consistency and accuracy in ratings.

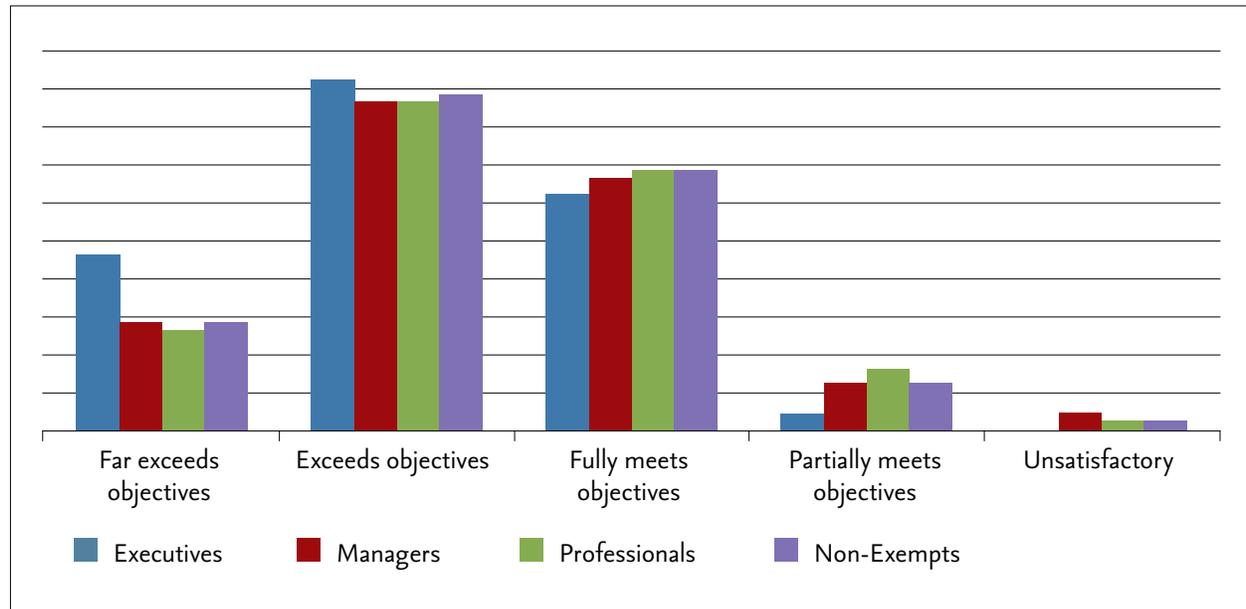
- ▶ *Number of Rating Levels.* Across sectors, the majority of organizations include from three to five

levels in their performance measurement systems. Performance can rarely be meaningfully assessed on a pass-fail basis, and it is not often useful to distinguish among more than five levels of performance.⁵¹ A 1994 survey of Fortune 100 companies found that most (60 percent) used appraisal systems with five levels of performance. Twenty percent of the surveyed organizations used more than five, and 20 percent used fewer than five.⁵²

There are advantages and disadvantages to each number of performance categories within the three-to-five range. Fewer categories may not allow for fine enough distinction between fundamentally different levels of performance, including outstanding performers or those who cannot improve. But fewer levels can foster greater consistency among raters. A larger number of performance categories make ratings more complex and tend to increase the chances that some levels will not be used at all. On the other hand, having more categories allows leaders to recognize truly excellent performance and to distinguish between low performers who can improve and those who should leave the organization. One particular advantage of an even number of performance categories (most often, four) is elimination of the frequently-used midpoint position, or the perception of “average” performance.⁵³ The most important quality of the performance categories is that they be clearly defined, so that expectations are clear among supervisors, ratings are more consistent, and employees understand the meaning of the designation they receive.⁵⁴

Even with a carefully selected number of performance categories, however, decades of research indicate that leniency and rating inflation are common problems across all types of industries.⁵⁵ For example, in the survey cited above, among the 80 percent of companies that used five or more levels of performance, raters actually used only the highest three levels (see Figure 4). Very few employees in these companies were rated in the bottom two performance levels. The norm, instead,

FIGURE 4. Performance Distribution among Fortune 100 Companies



Source: Based on data presented in Steven L. Thomas & Robert D. Bretz, Jr. “Research and Practice in Performance Appraisal: Evaluating Employee Performance in America’s Largest Companies.” *SAM Advanced Management Journal* (Spring 1994).

was a leniency bias, with most employees rated at the top end of the scale.⁵⁶

► **Category Descriptions.** Organizations can overcome some of the challenges outlined above by defining categories very specifically in terms of descriptively scaled behaviors and quantitatively defined categories. Quantitative categories can be clearly labeled with ranges, so that ratings are highly objective and subject to fact-checking. Qualitative measures also can be described clearly and sometimes transformed into quantitative, scaled categories. For example, rather than just having customers rate on a scale with such categories as “unsatisfied, satisfied, and very satisfied,” ratings can be more specific. On a rating of timeliness of checkout at a car rental service, the rating could have categories such as “under 3 minutes; 3–5 minutes; 5–10 minutes; and over 10 minutes,” making comparisons far more consistent. Details like these can be correlated with other measures so that the organization not only has information about individual employees, but

also can set clearer, more specific targets in the future.

Even the most qualitative measures like behavioral competencies can be broken down into scaled levels, with descriptions of behavior that distinguish high levels of the competency from middle and low, as shown in Figure 5. Scaled models enable organization leaders to classify the types of behavior they observe among their employees into meaningful, consistently scaled measures of performance, thus translating qualitative behaviors into quantified data.⁵⁷ Managers can then set standards for the level of rating on a given competency that is required for success.

Determining performance standards. By clearly identifying standards of excellence — both on individual measures and on employees’ overall performance — leaders help align both expectations and performance with the organization’s mission.⁵⁸ Across sectors, organizations set standards for employee performance in both the “what” and “how”

FIGURE 5. Sample Levels of Impact and Influence Competency

LEVEL	DESCRIPTION	BEHAVIORS
1	Personal Gain Tactics, Limited Persuasion	Uses negative behaviors for personal gain, or does not act to influence others.
2	No Adaptation to Audience	Prepares and presents data and logical arguments, but does not tailor to make them appealing or influential to the specific audience.
3	Tailors Single Action to Influence Audience	Thinks ahead about the likely reaction of audience, and adapts communication to obtain desired impact.
4	Tailors Single, Dramatic Action to Influence Audience	Takes one, dramatic action chosen to obtain a specific reaction from audience. (Threats do not count.)
5	Tailors Two Actions to Influence Audience	Thinks ahead about the likely reaction of audience, adapts communication to obtain desired impact, and shows “influence tenacity” by taking two (not necessarily dramatic) steps to influence.
6	Tailors Three Actions or Uses Indirect Influence	Takes three or more steps chosen to influence, or uses third-party experts or trusted individuals to influence others, or obtains individual support “behind the scenes,” or chooses timing and delivery/withholding of information to influence.
7	Complex Influence	Engages in a complex set of maneuvers with many people—personal communications, use of third parties, promotion decisions, sharing of power or information, working through chains of people for a “domino” influence effect—to obtain desired impact.

Source: Lyle Spencer and Signe Spencer, *Competence at Work* (New York: John Wiley, 1994).

categories by considering what is necessary to achieve the ultimate objectives and what is feasible within the constraints of the job. Key issues include deciding whether to use absolute or relative standards and how to account for changing circumstances.

Standards for output, or “what” measures, for example, might include selling 50 cases per week and establishing three new customers in the course of a month for a salesperson (see Figure 2 above). Other examples include the frequency expected of specific tasks and a specified time limit to complete a particular task. Standards for the “how” of an employee’s performance typically describe the type of skill or behavior expected in terms of its frequency, quality, or completeness. For example, an employer might set a target rating on customer satisfaction surveys, or identify target levels of behavioral competencies that are expected in a job or role.

In addition to defining performance standards

for individual objectives, organizations must also set standards for employees’ performance overall. These standards typically involve a roll-up of ratings on individual objectives, with some degree of weight assigned to each based on their importance to the employee’s and the organization’s success.⁵⁹

- ▶ *Using relative (versus absolute) targets.* The most straightforward method of evaluating employees is by judging their performance relative to a given standard, as described above.⁶⁰ Another approach to rating employees, however, evaluates their performance in comparison with one another rather than against a set standard.⁶¹ Performance ranking or “forced distribution” systems have been adopted among an increasing number of highly successful companies in large part to address the challenges of leniency and rating inflation, and to help inform decisions about compensation and

Performance ranking systems evaluate employees in comparison with one another rather than against a set standard, in part to address common challenges of leniency and rating inflation.

promotion.⁶² By some estimates, as many as 20 to 30 percent of all companies use performance ranking in some form.⁶³

Performance ranking typically involves assigning performance ratings to managerial and executive staff using forced distribution along a curve (e.g., 20 percent rated “most effective,” 70 percent rated “effective,” and 10 percent rated “needs improvement”). Some methods rank employees in a particular work group from best to worst using a “totem pole” approach. Other approaches place employees or teams into one of a set number of performance groups based on a fixed distribution (e.g., 25 percent into each of four groups).⁶⁴

General Electric is the best-known and most widely cited example of a successful forced ranking system. Former Chief Executive Jack Welch required managers to divide talent into three groups — the top 20 percent, middle 70 percent, and bottom 10 percent. The top-ranked performers were eligible for substantial performance awards, and many in the bottom 10 percent were shown the door.⁶⁵ In recent years, GE has softened its approach, continuing to use a curve for performance appraisal but using the 20-70-10 categories as general guidelines to inform pay and development decisions, rather than as required categories with default consequences.⁶⁶

To better recognize, reward, and retain its top talent in highly competitive Silicon Valley, internet search giant Yahoo recently instituted a totem pole approach to inform the allocation of salary increases and bonuses. Managers are asked to

rank employees within each unit and to distribute raises and bonuses accordingly. During reviews, they also discuss with employees how their raises compare with those of their colleagues.⁶⁷

Forced distribution systems are one way to overcome the pervasive tendency toward rating inflation and reinvigorate performance measurement that has become a rote exercise.⁶⁸ In addition, well-designed programs have been shown to be a highly effective way of dramatically increasing overall organizational results. In 2005, a group of university researchers constructed a sophisticated mathematical simulation of a forced ranking process across 100 companies. They found that forced ranking, including the firing of the bottom 10 percent of employees per year, resulted in an annual 16 percent gain in organization-wide productivity for the first two years.⁶⁹

Forced distribution programs have many critics, however, for their potential to negatively affect employee morale, teamwork and collaboration, future recruiting, and the opinions of external stakeholders.⁷⁰ Research suggests that the benefits of forced distribution drop off after the first two years, from 6 percent increases in productivity in the third and fourth years to essentially zero by year 10.⁷¹ Therefore, despite these systems’ strengths, many of even the most ardent supporters recommend implementing them for only a few years.⁷²

► *Accounting for idiosyncrasies.* In any work environment, there will inevitably be factors outside the organization’s and employees’ control that will nonetheless affect performance. For example, in the sales setting, a salesperson might meet her objective of selling 50 cases per week — and thus earn a satisfactory performance appraisal. But if, due to a shifting marketplace and consumer preferences, the salesperson’s market had actually expanded by 40 percent during the appraisal period, simply meeting this standard would not actually be satisfactory performance. Under the new circumstances, the firm might expect 70 cases per week instead.⁷³

Similarly, the character of a firm's neighborhood, the client base that it serves, or the degree of local market competition can change within a performance appraisal period and affect the results that an organization expects from its employees. Organizations across sectors rely on a variety of methods to take these changes into account when setting standards and evaluating employees' performance.⁷⁴ They do not stop measuring performance or setting standards. They simply change the measures and standards to fit new realities.

For example, in sales and many other settings, firms often account for changes and idiosyncrasies by benchmarking how other employees have achieved in similar situations, and setting standards for their own staff in comparison.⁷⁵ Similarly, some firms set standards for individual performance that automatically account for external conditions, such as changing competitive environments or socio-demographic characteristics. Using statistical modeling, these organizations estimate the results they expect for individuals or departments, which then serve as a point of comparison by which to evaluate actual performance.⁷⁶

Step Five: Adopt a Performance Measurement Process

With objectives and measures set for each employee, standards set, and a rating or ranking process in place, the key remaining design issue in a performance measurement system is the process by which performance reviews are conducted. Research from the fields of psychology, employee engagement, and performance management suggests that how performance appraisals are carried out is at least as important to employee and organization success as the performance measures upon which they are based.⁷⁷ In this section, we explore the key decisions for organization leaders in designing such a process, including determining who will carry out evaluations and how often.

Who conducts evaluations? Organizations take different approaches to the performance evaluation

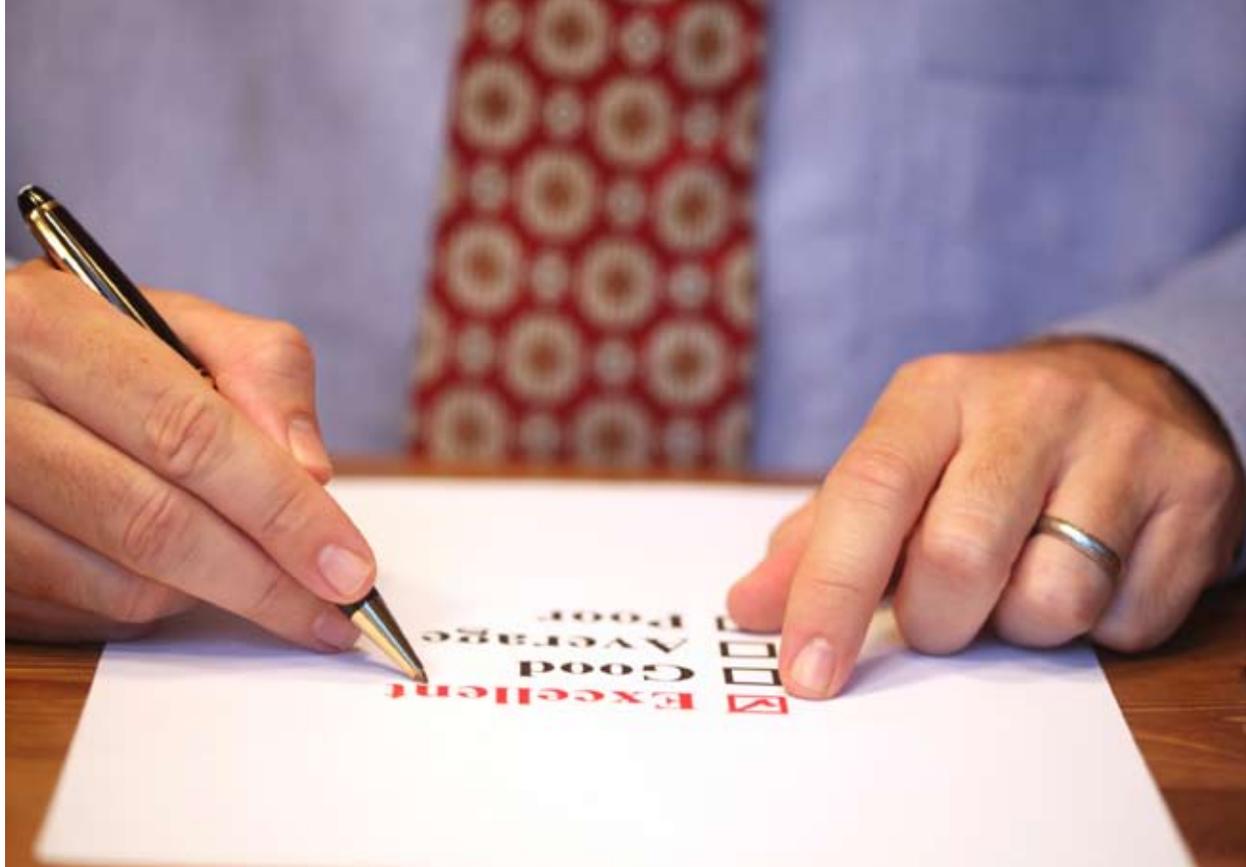
The process that leaders design to carry out performance appraisals is at least as important to success as the performance objectives they choose to assess.

process, with the majority assigning responsibility for appraisals to immediate supervisors. Many incorporate feedback from other players in the appraisal process, however, including peers, subordinates, customers, and employees themselves.

► *Reviews conducted by the direct supervisor.* In the great majority of organizations, the employee's direct supervisor oversees performance appraisals, including collecting hard data, evaluating the employee's competencies and progress, and conducting the appraisal discussion.⁷⁸ This is largely because the direct supervisor typically has the most direct contact with and authority over the employee, and therefore is likely to best understand the nature of the employee's work and overall performance.⁷⁹

While supervisors are the most common evaluators and have clear advantages in this role, research and experience suggest some potential shortcomings to their ratings. For example, research indicates that manager evaluations can be unreliable for a number of reasons, including the manager's desire to improve perceptions of performance within his or her own workgroup, lack of time or energy to dedicate to the task, and misunderstandings about the process or goals of performance measurement.⁸⁰ In addition, supervisory ratings have been shown to contribute to the leniency and rating bias described above, as managers seek to avoid uncomfortable confrontations, encourage an employee to improve, or move an unsuccessful employee into a new work setting.⁸¹

Research and experience from other sectors



suggest two methods for overcoming shortcomings of supervisors' appraisals. First, holding managers accountable for the quality and substance of their appraisals can improve the reliability and meaningfulness of the performance measurement process.⁸² In an extensive review of evaluation practices among Fortune 100 companies, researchers found that less than 25 percent of managers were evaluated on how they conduct performance

encourage them to assign more accurate ratings and invest more in the process.⁸³

Providing detailed and frequent training is another tactic some experts support to produce meaningful, fair, and consistent evaluations, given the complexity of the decisions managers must make during the performance measurement process.⁸⁴ But in fact, evaluators typically receive limited training on evaluation processes, with little follow-up beyond initial training sessions focused on providing feedback, using appraisal forms, or avoiding rating errors.⁸⁵ For example, a 1991 survey of 22 civil engineering firms found that less than 40 percent of reviewers received training in the appraisal process.⁸⁶

► *Reviews conducted by other stakeholders.* Many organizations across sectors include others in the evaluation process in addition to the direct supervisor, including higher-level supervisors, peers, subordinates, and customers. These additional perspectives can reduce the likelihood of bias in the evaluation process and provide a broader picture of the employee's performance.⁸⁷

By holding managers accountable for the quality and substance of their appraisals, organization leaders can improve the reliability and meaningfulness of the performance measurement process.

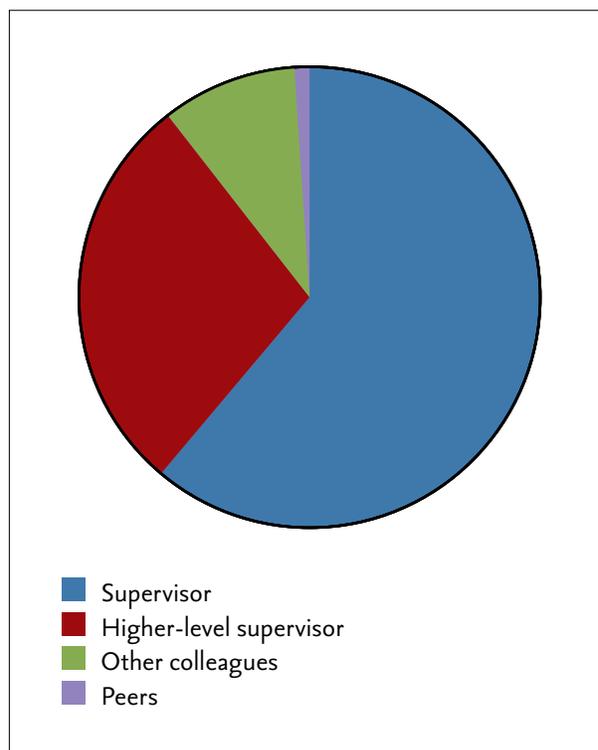
appraisals or award performance ratings. By incorporating the outcomes of appraisals into the manager's own evaluations, organization leaders can

One popular method for incorporating feedback from a broader range of stakeholders is the 360-degree review, which includes appraisals from managers, coworkers, subordinates, and the employee to provide a “360-degree” picture of the employee’s performance.⁸⁸ Many organizations have found these reviews to be a critical element of their performance measurement process, both for providing a fuller view of performance and increasing employees’ buy-in into the process.⁸⁹ Microsoft, for example, asks all employees to complete appraisal forms about their manager’s performance on an annual basis. Questions cover the degree to which the manager empowers the employee to do his job, provides an environment that motivates employees to achieve their goals, and recognizes and rewards information and creativity.⁹⁰ This type of feedback is particularly helpful for employee (and manager) development, by identifying areas of strength and weakness to inform future training or development opportunities.⁹¹ At Microsoft, the process also serves to reinforce core values throughout all levels of the organization.

It is not clear from the research, however, that multisource appraisals increase the validity of final performance ratings, or that they should be used for purposes beyond employee development.⁹² Other colleagues and subordinates often suffer from the same leniency and rating bias as direct supervisors, frequently inflating ratings to avoid confrontations or increase the likelihood that the employee will assign them a high performance rating in return.⁹³

When multisource feedback is used in final performance ratings, reviews from those other than the supervisor are typically weighted less heavily in the total performance rating. In their survey of Fortune 100 companies, Thomas and Bretz found that the rating from an employee’s immediate supervisor is often weighted most heavily (comprising between 42 and 74 percent of the total performance rating) followed by the

FIGURE 6. Relative Importance of Reviewer Feedback among Fortune 100 Companies



Source: Based on data presented in Steven L. Thomas & Robert D. Bretz, Jr. “Research and practice in performance appraisal: evaluating employee performance in America’s largest companies.” *SAM Advanced Management Journal* (Spring 1994).

higher-level supervisor (18 to 36 percent), other colleagues (3 to 15 percent) and peers (0 to 2 percent) (see Figure 6).⁹⁴ Former General Electric CEO Jack Welch, a vocal proponent of 360-degree reviews, argues that the process is primarily helpful when used in the course of evaluations every second or third year, to provide an opportunity for otherwise silent colleagues to offer their perspective.⁹⁵

- *Reviews conducted by the employee.* Some companies allow employees to offer a self-evaluation as part of the overall appraisal process. Self-reviews can provide employees with an opportunity to examine their strengths and weaknesses, and to discuss areas for growth with a supervisor. In addition, it has been suggested that evaluating one’s

own effectiveness can enhance the assessment process by increasing buy-in in the process and impact of review feedback.⁹⁶

Several studies conducted under a variety of research conditions show that self-appraisals are especially subject to leniency bias, however, and therefore have low validity for objectively assessing performance.⁹⁷ In the survey of Fortune 100 companies cited above, companies did not typically use ratings resulting from self-appraisals in the formal evaluation process, or they assigned them very little weight (about 1 percent) in an employee's final rating.⁹⁸

How frequent are performance appraisals? Across sectors, most organizations conduct formal performance evaluations once a year. A 2010 survey of more than 500 human resources managers in large and midsize companies found that 60 percent conducted performance appraisals annually.⁹⁹ These typically consist of performance feedback, including discussion of the employee's strengths and weaknesses, as well as management decisions about issues such as compensation, promotions, or training.¹⁰⁰

Many experts advocate for performance measurement systems that incorporate more than one appraisal per year, however, particularly for purposes of employee feedback and development. In a survey of managers in 22 civil engineering firms, researchers identified evaluating employees at least twice a year as a "strong point" of the performance appraisal process.¹⁰¹ A 2008 survey of accountants working in government, industry, and public accounting found that the majority of respondents would prefer one more evaluation than they had, for a total, most commonly, of two per year.¹⁰² Former GE CEO Jack Welch also argues for "rigorous and candid" performance evaluations at least semiannually to give employees an opportunity to know where they stand relative to the organization's standards and their colleagues.¹⁰³

The research evidence does not suggest an ideal frequency of performance evaluations for any given employee or organization. Instead, experience sug-

gests that multiple criteria should guide the frequency of performance appraisals, including:

- ▶ *The availability of new and meaningful employee data.* New and meaningful data are not always available to warrant performance discussions with employees more than once or twice per year.¹⁰⁴ Formal evaluations based on interim collections of data could cause both the manager and the employee to direct their efforts in shortsighted ways or to reward or sanction only short-term performance. On the other hand, sometimes new and critical data become available to warrant a formal appraisal, and the ideal performance measurement system will allow sufficient flexibility for these discussions to occur.¹⁰⁵
- ▶ *How the appraisal will be used.* Decisions about the frequency of performance appraisals should also be informed by the purposes for which they will be used. For example, many human resource experts suggest that evaluations used for development purposes should occur more often than summative assessments used primarily for pay or promotion. More frequent feedback on employees' strengths and challenges can enable them to more quickly redirect their behaviors and pursue more rapid growth.¹⁰⁶
- ▶ *The benefit of the result relative to its costs.* Performance evaluations consume valuable resources, both in terms of managers' and employees' time and the organization's resources. While evidence suggests that most managers do not spend enough time conducting thorough and meaningful performance appraisals, researchers estimate that the average manager still devotes between four and eight hours per employee to the process over the course of the year.¹⁰⁷ In addition, the development of quantitative and qualitative performance measures and tools throughout the process requires up-front investment from an organization and regular updating and revision. Therefore, appraisals should be used as often as organization leaders believe that benefits justify their costs.¹⁰⁸

Step Six: Use Measurement Results to Take Action

One of the primary goals of performance measurement is to enable organization leaders to make fair, informed, and meaningful management decisions that increase the effectiveness of the organization. A strong appraisal system will arm organization leaders with data to make good decisions about development, rewards (pay increases and bonuses), career opportunities and advancement, and future recruiting and selection.

In addition, a well-documented process can help protect the firm with records of performance when dismissals are necessary. Each of these areas warrants its own report of best practices. In this section, we briefly discuss the prevalence and value of each use of performance measurement data and provide examples of how they play out in organizations across sectors.

► **Development.** A key purpose of performance appraisals, as suggested under Step One, is to provide feedback to employees about their performance and further develop their capacity to perform well.¹⁰⁹ Survey evidence suggests that, across sectors, this is among the most common uses of performance evaluations.¹¹⁰

For example, at Public Service Electric & Gas, one of the largest investor-owned utilities in the country, managers used performance appraisals to initiate discussions with employees about specific short- and long-term areas for development. They also found, however, that as a byproduct of the process, because employees had more regular insight into their performance, they were able to understand the causes of certain challenges on their own. Instead of speculating about why customers weren't satisfied with repair services, for example, workers and supervisors could look at consumer satisfaction metrics, see the reasons for those complaints, and make a point of improving particular aspects of the process in the next quarter.¹¹¹

Appraisal conversations that focus on growth

A strong performance measurement system arms leaders with data to make good decisions about employees' development, rewards, career opportunities and dismissal — as well as to inform the organization's future recruiting and selection.



and development offer a number of benefits for employees and the organization as a whole, including providing meaningful opportunities for employee engagement, identifying common training and development needs across the firm, and setting future objectives to improve overall success.¹¹²

Cross-sector research and experience suggest that three key ingredients of a performance measurement system can make the most of these dis-

cussions. Primary among these is for the appraisal to provide meaningful information about how the employee is performing, including specific areas for improvement and the basis for concrete next steps to develop necessary skills or enhance existing ones.¹¹³ To ensure that this type of information is included in the final appraisal, evaluations should include multiple sources of qualitative data through methods such as 360-degree reviews or individual development reviews with a supervisor.¹¹⁴ The appraisal process should also provide an opportunity for the employee to understand what level of improvement is necessary to obtain rewards or recognition, such as promotions or salary increases, and the time period in which improvements will be expected.¹¹⁵

► **Rewards.** Along with employee development, administering salaries and financial rewards is the most common purpose of performance appraisals.¹¹⁶ In their survey of Fortune 100 companies, Thomas and Bretz found that 92 percent used information from performance appraisals to determine merit pay increases for employees.¹¹⁷ At Citibank, for example, branch managers' final performance evaluations were linked to annual bonus determinations under a very straightforward system. Managers who received a rating "below par" received only standard base pay. A "par" rating generated a bonus of up to 15 percent of the base salary, and an "above par" rating could lead to a bonus as high as 30 percent of the manager's base pay.¹¹⁸ Surveys of high-performing governmental organizations suggest that most also link performance appraisals in some way to employees' pay.¹¹⁹ In addition, many organizations that incorporate forced distribution into their performance measurement systems rely upon rankings of employees to determine bonuses and pay increases.¹²⁰

Research and experience from across sectors suggest that to inform decisions about pay, performance measurement systems should incorporate three key features. First, as outlined under Step Two, appraisals must be based upon fair measures

of performance and focus on all aspects of the job that are important to the employee's success.¹²¹ They should also be paired with frequent feedback on progress.¹²² As a technical matter, however, actual conversations about pay increases should be conducted separately from those about the employee's development, if possible. Research is mixed on the advantages of combining or separating the two, but generally suggests that when the two are combined, reflection on past performance and the amount of the performance award can eclipse important conversations about opportunities for future growth and development.¹²³ As management expert Dick Grote explains, "concentrating the performance appraisal discussion on the appraisal itself and assigning the compensation discussion to a separate meeting can ensure maximum mileage from both."¹²⁴

► **Dismissal.** Organization leaders also rely upon performance appraisals to make management decisions about demotion and dismissal.¹²⁵ When organization leaders must make these decisions, strong performance measurement systems can provide the necessary record and justification to support termination or demotion, and protect the organization and its managers should the decision be challenged in court.¹²⁶ Of course, no performance measurement system can protect an organization from unwarranted terminations or those based on illegal grounds such as the employee's race, age, sex, religion, or other protected characteristics — nor should they. But when an employee's performance warrants removal or other similar action, performance measurement systems that include frequent evaluation and documentation of an employee's work, fair measures of performance, and reasonable opportunity for the employee to improve can help uphold the organization's decision.

Chinese appliance manufacturer Haier, widely recognized as having one of the world's strongest performance management systems, regularly uses data from performance appraisals to inform

dismissal decisions. Under its forced distribution program, Haier annually dismisses the lowest 10 percent of employees through a three-phase process. In the first annual or quarterly review in which an employee is in the bottom 10 percent, he or she is put “on leave” and sent for job training at Haier’s expense. If he or she remains in the bottom 10 percent during the second review, the employee is required to attend a second round of job training, but this time at his or her own expense. Thereafter, continued performance in the bottom 10 percent results in automatic dismissal.¹²⁷

► **Future Recruiting and Selection.** Organizations can also use strong performance evaluations to correlate objective results data with the behaviors, competencies, and skills of an organization’s highest performers — that is, to analyze and identify relationships between the “what” and “how” objectives outlined in Step Two. Because competencies can also be evaluated before a candidate actually performs on the job, information about the behaviors and skills of high performers enables organizations to recruit and screen new candidates based on characteristics that have been shown to matter in a particular job. Indeed, hiring processes based on rigorous assessment of the competencies that correlate with strong performance among current employees are one of the most cost-effective strategies for improving the accuracy of selection.¹²⁸ In addition, when a strong performance measurement system enables tying performance to pay, research indicates that higher-quality and higher-performing employees disproportionately choose to work with the organization in the future.¹²⁹

In 2009, researchers used this method to validate and refine performance evaluations and recruiting in the British Royal Navy. They used a competency-based interview method to correlate behaviors and skills of officers with the Navy’s own appraisals of officers’ performance. The results showed four competency clusters associated with high performance, which the Navy thereaf-

ter adopted as the basis for all screening and selection of leadership and management candidates.¹³⁰ Many other private sector organizations that incorporate competencies into their performance measurement process are similarly able to correlate competencies with high performance and use the results to guide future recruiting and selection.

Research and evidence from across sectors show that a strong performance measurement system enables organization leaders to make critical decisions about employees, strategies, and firm-wide success. By considering each of these six critical steps, leaders in a variety of organizational settings have built strong performance measurement systems to foster excellence and continuous improvement among their employees and increase overall organization success.



Applying Cross-Sector Lessons to Performance Measurement of Educators

The research and experience base on performance measurement across sectors is so rich in large part because organization leaders in other industries know that talent — and strategic management of that talent — is the key to their competitiveness and success.

Yet there is no sector where talent is more important than education. We have a growing body of research showing that teachers' and principals' contributions to student learning vary widely, and the differences between the best and worst educators dramatically alter students' lives and future chances of success. Measuring educators' performance accurately — and in ways that enable crucial decisions — must therefore be a top priority within education as it is in other sectors. Improving performance measurement for teachers and principals is not just a step but the very foundation of refocusing our public education system on achieving student outcomes.

Fortunately, education leaders have a great deal to learn from many other sectors and organizations that have led the nation in performance measurement. These sectors have dealt with many of the same challenges facing education leaders, and they have devised, studied, and time-tested solutions to address those challenges. We can use these lessons to inform performance evaluation reforms for teachers and principals, such as developing objectives, measures, and standards according to the best practices across sectors, and using them to guide management decisions in education.

Figure 7 draws from the cross-sector lessons to present an example of a performance measurement system for teachers. Table 1 offers an example of how

the levels and behaviors of a specific competency can be displayed for performance measurement. These are not meant to be prescriptive but are simple illustrations of how the lessons about performance measurement from other sectors can be applied in the education setting.

In addition to these sample models, we offer a few **additional recommendations for education leaders committed to improving performance measurement** for teachers and principals:

- ▶ **Measure more to measure better.** Too often in education, our fear of imperfect measures trumps the will to improve evaluation systems at all. Districts, schools, and other providers that are committed to improving the performance of professionals who enter and stay in education — and how much they contribute on the job — will measure fearlessly and frequently, and use the shortcomings of current systems to spur constant improvement. With a simultaneous commitment to use the lessons and measurement tools available today and improve them in the future, we can rapidly advance performance measurement — and contribute to improved student outcomes.
- ▶ **Measure both the “what” and “how” of educators' performance.** Because the education sector has such limited evidence about the process skills and competencies that contribute to outstanding work, many have become distracted by controversies about evaluations that are based only on student results. But the research outlined above reveals the clear advantages to including measures in evaluations of student results *and* educators' skills and competencies. Combining both kinds

FIGURE 7. Sample Elements of a Performance Measurement System for Teachers

Goal	Objectives	Measures	Standards	Process	Uses
Prepare students for success in college, career, and life	Achieve significant learning gains for all students	Student growth on state assessments (all students)	Growth rate for all students, as defined by other teachers' rates in previous years	Principal Role: evaluation twice per year (using student work review and students' ratings of teacher's competencies midyear; state assessment results and peer and observer competency ratings at year's end)	<ul style="list-style-type: none"> • Development • Tenure review • Promotion • Compensation
		Student growth on state assessments (subgroups)	Growth rate for all student subgroups, as defined by other teachers' rates w/ similar students		
		Student achievement on state assessments (% at grade level)	State proficiency targets		
		Progress on ratings of student work (using standardized rubrics)	Meet rating targets and progress goals on each rubric element		
	Demonstrate competencies correlated with positive student outcomes	Ratings of competency levels by trained observers (see Table 1)	Threshold or superior levels of competence correlated with target student outcomes		Development Career paths Selection

Note: Blue boxes describe the objectives, measures and standards related to the “what” of a teacher’s job, while white boxes describe the objectives, measures, and standards related to the “how” of their success. Tan boxes apply to both categories.

of measures will increase the developmental value of evaluations and lay the groundwork for better-informed decisions about compensation, promotion, and dismissal.

- ▶ **Conquer fears of measuring student learning.** While measures of skills and competencies have a critical role in educators’ evaluations, district and state leaders must also consider data about

teachers’ and leaders’ contributions to student progress. After all, schools are in the business of learning — and students’ academic growth is the primary outcome. Rather than fear student learning results, we must measure them — and commit to improve our measurement over time, rather than using imperfections as an excuse to resist such measures entirely.

TABLE 1. Sample Levels of Achievement Competency for a Teacher

Zone	Level	Description	Behaviors
Red Flag	1	Low concern for work quality	Shows little concern for quality of work, or preoccupied by non-work matters
Neutral	2	Wishes to do job well	Expresses desire to do the job well but does not make measurable improvements or have a clear standard of excellence
Threshold	3	Moderate concern for work or quality	Works to do tasks and meet standards required by principal/management or makes voluntary improvements, but with no specific goal in mind, or to meet only modest goals
	4	Strong concern for work goals and quality improvement	Sets challenging but realistic work goals for self, students, or other adults and acts to meet them, or sets challenging goals for self and students and monitors progress
Superior	5	Prioritizes goals and tasks based on impact relative to effort	Carefully chooses challenging goals and actions towards goals (for self and students) based on cost-benefit analysis: time, money, and other resources needed, versus the speed and magnitude of results
	6	Pursues high-risk goals and improvement	Commits significant resources and time to reach a very challenging goal without being sure of success, and takes multiple actions to minimize risk (e.g., conducting research, anticipating barriers, planning ahead, engaging others to help)

Source: Public Impact (2008). *School Turnaround Teachers: Selection Toolkit*, with competencies derived from Spencer & Spencer (1993). *Competence at Work*; Hobby, Crabtree and Ibbetson (2004). *The School Recruitment Handbook, A Guide to Attracting, Selecting and Keeping Outstanding Teachers*; and Haberman (1995). *Star Teachers of Children in Poverty*.

► **If a measure does not predict success, change it.**

Great organizations change their measures, and measurement processes, to improve their validity and to reflect changes in how work is done. Education must do no less.

► **Meaningfully assess team performance.** Many employees' contributions occur largely at the team level, and successful organizations across sectors have developed careful strategies to assess their performance in that context. Methods include identifying individual contributions to team results using peer evaluation and statistical analysis, and assessing team outcomes alongside individual results.

► **Adopt strategies to protect against leniency and bias.** Tactics include rating on a forced distribution, holding managers accountable for the appraisals they conduct, including reviews from multiple perspectives with different weights, and providing training in the performance measure-

ment process. Each of these strategies — including forced distribution systems in particular — must be designed carefully to offer their full benefit. But the alternatives typically suffer from leniency that robs educators of valuable feedback and students of the teachers they deserve.

► **Use performance measurement results.** Improvements to our performance measurement systems will have no effect unless the results are put to use. States, districts, and other providers must commit to act upon real differences in educator effectiveness — by using evaluations as the basis for teachers' and principals' ongoing development and pay, decisions about retention and dismissal, and future recruiting and selection. Altogether, these decisions form the culture in our schools — the culture that determines who decides to teach, how long they stay, what students they reach, and how motivated they feel to raise our children's sights and conquer obstacles.

Endnotes

1. See, e.g., The New Teacher Project, *Teacher Evaluation 2.0*. (NY: The New Teacher Project, 2010); L. Darling-Hammond. *Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching*. (Washington, DC: Center for American Progress, 2010).
2. See, e.g., D. Weisberg, S. Sexton, J. Mulhern, & D. Keeling, *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness* (NY: The New Teacher Project, 2007).
3. R.D. Behn, "Why Measure Performance: Different Purposes Require Different Measures," *Harvard Business Review* (2003); Satoris S. Youngcourt, Pedro I. Leiva, & Robert G. Jones, "Perceived Purposes of Performance Appraisal: Correlates of Individual- and Position-Focused Purposes on Attitudinal Outcomes," *Human Resource Development Quarterly*, 18,3 (Fall 2007), 315–343; Chen-Mind Chu & Dar-Hsin Chen, "Performance Appraisal Systems in Service and Manufacturing Industries: Evidence from Taiwan," *International Journal of Management*, 4, 3 (September 2007), 512–521; J.N. Cleveland, K.R. Murphy, & R.E. Williams, "Multiple Uses of Performance Appraisal: Prevalence and Correlates," *Journal of Applied Psychology* 74, 1 (1989), 130–135.
4. Kevin J. Williams, "The Analytic and Non-Analytic Frameworks for Performance Appraisal," *Human Resources Management Review*, 2,1 (1992), 1–8; B.N. Smith, J.S. Hornsby, & R. Shirmeyer, "Current Trends in Performance Appraisal: An Examination of Managerial Practice," *Advancement Management Journal* (Summer 1996); Cleveland et al. (1989).
5. T. Gallagher, "360-Degree Performance Reviews Offer Valuable Perspectives," *Financial Executive*, 24,10 (December 2008), p. 61; W.R. Boswell & J.W. Boudreau, "Separating the Developmental and Evaluative Performance Appraisal Uses," *Journal of Business and Psychology*, 16, 3 (Spring 2002), 391–412.
6. Steven L. Thomas & Robert D. Bretz, Jr. "Research and Practice in Performance Appraisal: Evaluating Employee Performance in America's Largest Companies," *SAM Advanced Management Journal*, 59 (Spring 1994); Linda S. Pettijohn, R. Stephen Parker, Charles E. Pettijohn, & John L. Kent, "Performance Appraisals: Usage, Criteria, and Observation," *The Journal of Management Development*, 20, 9/10 (2001), 754–770; Cleveland et al. (1989).
7. D.P. Bozeman, "Interrater Agreement in a Multi-Source Performance Appraisal: A Commentary," *Journal of Organizational Behavior*, 18 (1997), 313–316; J. Grote, *The Complete Guide to Performance Appraisal*, (New York: AMACOM, 1996); J. Dean, "Managing," *Fortune Magazine*, (December 1993); M. London & J.W. Smither, "Can Multi-Source Feedback Change Self-Evaluation, Skill Development and Performance?" *Personnel Psychology*, 48 (1995), 375–390; J.F. Milliman, R.A. Zawacki, C. Normal, L. Powell, & J. Kirksey, "Companies Evaluate Employees from All Perspectives," *Personnel Journal* (November 1994), 99–103.
8. Behn (2003); H. Levinson, "Appraisal of What Performance?" *Harvard Business Review*, 54 (1976); C. Moon, J. Lee, C. Jeong, J.E. Sungcheol, & P.S. Lim, "An Implementation Case for Performance Appraisal and Promotion Ranking," IEEE International Conference on Systems, Man, and Cybernetics, ISIC (July 10, 2007).
9. T.P. Flannery, D.A. Hofrichter, & P.E. Platten, *People, Performance and Pay: Dynamic Compensation for Changing Organizations*, Hay Group (1996); Grote (1996); Behn (2003); Levinson (1976).
10. L.S. Pettijohn, R.S. Parker, C.E. Pettijohn, & J.L. Kent, "Performance Appraisals: Usage, Criteria, and Observation," *The Journal of Management Development*, 20, 9/10 (2001), 754–770; Cleveland et al. (1989).
11. V. Dulewicz & M. Young, "A Study into Leadership and Management Competencies Predicting Superior Performance in the British Royal Navy," *Journal of Management Development*, 28, 9 (September 2009), 794–820; D. Harrison, "Best Practices in Assessments Drive Performance, Development and Retention," Impact Achievement Group (2008); L.M. Spencer, D.C. McClelland, & S.M. Spencer, *Competency Assessment Methods: History and State of the Art* (Hay/McBer Research Press, 1997).
12. See., e.g., T.R. Zenger, "Why Do Employers Only Reward Extreme Performance? Examining the Relationships among Performance, Pay, and Turnover," *Administrative Science Quarterly*, 37, (1992), 198–219; Rajiv D. Banker, Seok-Young Lee, & Gordon Potter, "A Field Study of the Impact of a Performance-Based Incentive Plan," *Journal of Accounting and Economics*, 21 (1996), 195–226; Jan Bouwens & Laurence van Lent, "Effort and Selection Effects of Incentive Contracts," (Tilburg, the Netherlands: Tilburg University, 2003); Michael Beer & Mark D. Cannon, "Promise and Peril in Implementing Pay-For-Performance," *Human Resource Management*, 43, 1 (Spring 2004), 3–48; Christopher M. Lowery, N. A. Beadles II, M. M. Petty, Gordon M. Amsler, & James W. Thompson, "An Empirical Examination of a Merit Bonus Plan," *Journal of Managerial Issues*,

- 14, 1 (Spring 2002), 100–117; Hay Group, *Bridging the Pay-for-Performance Gap: Establishing Truly Differentiated Rewards* (Philadelphia, PA, 2004).
13. Pettijohn et al. (2001).
14. M. Bourne & A. Neely, “Implementing Performance Management Systems: A Literature Review,” *International Journal of Business Performance Management*, 5, 1 (2003), 2–24.
15. Braam Verster, Hannah Wozniak, Nigel Manson, Richard Dobbs, & Toby Gibbs, *Performance Management: Case Studies of World-Class Operators*, (Washington, DC: The Aspen Institute, 2007).
16. Bourne & Neely (2003); The New Teacher Project, *Performance Measurement*, (Washington, DC: The New Teacher Project, 2010).
17. Grote (1996); Levinson (1976); Hay Group, “View Point: Performing in Uncertain Times,” June 2010, Issue 3; R.K. Thomas, P. Cheese, & J.M. Benton, “Human Capital Development,” *Accenture Research Notes*, Issue One (Nov. 1, 2003).
18. Hay Group (2010).
19. Grote (1996); L. Spencer & S. Spencer, *Competence at Work* (New York: John Wiley, 1994).
20. Grote (1996); T.P. Flannery, D.A. Hofrichter, & P.E. Platten, *People, Performance and Pay: Dynamic Compensation for Changing Organizations* (New York: Hay Group, 1996); R. Kunreuther, “Federal Employee Performance Appraisal: The Good, Bad, and Ugly,” FedSmith.com (Oct. 7, 2009).
21. Thomas & Bretz (1994).
22. G. Roberts & T. Reed, “Performance Appraisal Participation, Goal Setting and Feedback.” *Review of Public Personnel Administration*, 30 (Fall 1996), 29–60; A. Kluger & A. DeNisi, “The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory,” *Psychological Bulletin*, 119, 2 (1996), 254–284.
23. Thomas & Bretz (1994).
24. Verster et al. (2007); Thomas & Bretz (1994); G.R. Gilbert & A.E. Nelson, “The Pacer Share Demonstration Project: Implications for Organizational Management and Performance Evaluation,” *Public Personnel Management*, 18, 2 (Summer 1989), 209–225; I.M. Jawahar, “Correlates of Satisfaction with Performance Appraisal Feedback,” *Journal of Labor Research*, 27, 2 (2006), 213–236; Youngcourt et al. (2006); R.D. Arvey & K.R. Murphy, “Performance Evaluation in Work Settings,” *Annual Review of Psychology*, 49 (1998), 141–68.
25. N. Nohria, B. Groysberg, & L.E. Lee, “Employee Motivation: A Powerful New Model,” *Harvard Business Review* (July-August 2008); Jawahar (2006); Verster et al. (2007); Youngcourt (2004); F. R. Reichheld & P. Roger, “Motivating through Metrics,” *Harvard Business Review* (2005); D. Fandray, “The New Thinking in Performance Appraisals,” *Workforce Magazine* (May 2001).
26. Jawahar, 2006; T.S. Clausen, K.T. Jones, & J.S. Rich, “Appraising Employee Performance Evaluation Systems,” *CPA Journal* (February 2008).
27. L. Anderson & D. Cochenour, “Merit Salary Criteria: One Academic Library’s Experience,” *Libraries and the Academy*, 1, 4 (October 2001), 465–480; R.S. Norton & D.P. Kaplan, “Using the Balanced Scorecard as a Strategic Management System,” *Harvard Business Review* (1996); R.G. Eccles, “The Performance Measurement Manifesto,” *Harvard Business Review* (January-February 1991).
28. Grote (1996); M.T. Strebler, S. Bevan, D. Robinson, “Performance Review: Balancing Objectives and Content,” *Institutes for Employment Studies* (2001); H.H. Meyer, E. Kay, & J.P. French, “Split Roles in Performance Appraisal,” *Harvard Business Review* 44, 3 (May/June 1966).
29. G.B. Sprinkle & M.G. Williamson, “The Evolution from Taylorism to Employee Gainsharing: A Case Study Examining John Deere’s Continuous Improvement Pay Plan,” *Issues in Accounting Education*, 19, 4 (2004), 487.
30. Thomas & Bretz (Spring 1994); see also Hay Group (2010).
31. D. Everatt, V. Pucik, & K. Xin, “Managing Performance at Haier,” *International Institute for Management Development* (2004); S. Rosegrant, R. Shapiro, and M. Watkins, “A Measure of Delight: The Pursuit of Quality at AT&T Universal Card Services,” Harvard Business School Case Study (1997).
32. Norton & Kaplan (1996); Hay Group (June 2010).
33. J.B. Shah & J. Murphy, “Performance Appraisals for Improved Productivity,” *Journal of Management in Engineering*, (March/April 1995) 26–28; Cleveland et al. (1989); Eccles (1991); Banker et al. (1996).
34. Shah & Murphy (1995); Cleveland et al. (1989); Eccles (1991); Banker et al. (1996).
35. A. Dávila & R. Simons, *Citibank: Performance Evaluation*, Harvard Business School Case Study (October 1999).
36. Reichheld & Rogers (2005).
37. Spencer & Spencer (1994).
38. Behn (2003); Cleveland et al. (1989); R. Kunreuther, “Federal Employee Performance Appraisal: The Good, Bad, and Ugly,” FedSmith.com (Oct. 7, 2009).
39. Oak Enterprises, *Understanding IT Worker Productivity and Finding Those Valuable “2x, 4x, 10x” Performers*, (Glen Ellen, IL: Oak Enterprises, n.d.); Flannery et al. (1996).
40. Grote (1996).

41. Grote (1996); Spencer & Spencer (1994).
42. Grote (1996); Spencer & Spencer (1994).
43. J.A. Drexler, T.A. Beehr & T.A. Stetz, "Peer Appraisals: Differentiation of Individual Performance On Group Tasks," *Human Resource Management*, 40, 4 (Winter 2001), 333–345; J.S. Miller & R.L. Cardy, "Self-Monitoring and Performance Appraisal: Rating Outcomes in Project Teams," *Journal of Organizational Behavior*, 21 (2000), 609–626.
44. S.G. Scott & W.D. Einstein, "Strategic Performance Appraisal in Team-Based Organizations: One Size Does Not Fit All," *Academy of Management Executive*, 15, 2 (2001); Grote (1996).
45. Grote (1996).
46. D.J. Weiss. "Extracting Individual Contributions to a Team's Performance." *Teorie & Modelli* (2007) 9: 1–2.
47. Grote (1996).
48. Scott & Einstein (2001).
49. Miller & Cardy (2000); Scott & Einstein (2001).
50. Scott & Einstein (2001).
51. Grote (1996).
52. Thomas & Bretz (1994).
53. Grote (1996).
54. X. M. Wang, K.F. Wong, J.Y. Kwong, "The Roles of Rater Goals and Ratee Performance Levels in the Distortion of Performance Ratings," *Journal of Applied Psychology*, 95, 3 (2010), 546–561; S.A. Way, M.C. Sturman, & C. Raab, "What Matters More?" *Cornell Hospital Quarterly*, 51, 3 (2010), 397.
55. Wang et al. (2010); Yithak Fried, Ariel S. Levi, Haim Ailan Ben-David, & Robert Tieg, "Inflation of Subordinates' Performance Ratings: Main and Interactive Effects of Rater Negative Affectivity, Documentation of Work Behavior, and Appraisal Visibility," *Journal of Organizational Behavior*, 20 (1999), 431–444; Arvey & Murphy (1998); Grote (1996); C.O. Longnecker, H.P. Sims & D.A. Gioia, "Behind the Mask: the Politics of Employee Appraisal," *The Academy of Management Executive* 1,3 (1987), 183–193.
56. Thomas & Bretz (1994).
57. Spencer & Spencer (1994).
58. Levinson (1976); Grote (1996); G. Baker, R. Gibbons & K.J. Murphy. "Subjective Performance Measures in Optimal Incentive Contracts." *The Quarterly Journal of Economics* (1994) 109,4: 1125–1156; F.A. Muckler, "Selecting Performance Measures: 'Objective' versus 'Subjective' Measurement," *Human Factors* 34, 4 (1992), 441–455.
59. Thomas & Bretz (1994); R.D. Bretz, G.T. Milkovich, & W. Read, "The Current State of Performance Appraisal Research and Practice: Concerns, Directions, and Implications," *Journal of Management* 18 (1992), 321.
60. Mostafa Jafari, Atieh Bourouni, & Roozbeh Hesam Amiri, "A New Framework for Selection of the Best Performance Appraisal Method," *European Journal of Social Sciences*, 7, 3 (2009).
61. R. Grote, "Forced Ranking: Making Performance Measurement Work." *Harvard Business Review*, (November 2005). Many organizations use forced distributions or ranking, particularly in conjunction with pay decisions, to supplement their main performance appraisal method, especially for managers and professional employees. See also Thomas & Bretz (1994).
62. Grote (2005); C.A. Olson & G.M. Davis, "The Pros and Cons of Forced Ranking." *Society for Human Resources Management* (2003).
63. L. Gary, "The Controversial Practice of Forced Ranking," *Harvard Management Update* (June 2001); Grote (2005).
64. Olson & Davis (2003).
65. J. McGregor, "The Struggle to Measure Performance," *Business Week* (January 2006); Grote (1996).
66. D. Brady, "Secrets of an HR Superstar," *Business Week* (April 2007).
67. McGregor (2006).
68. Olson & Davis (2003).
69. Steven E. Scullen, Paul K. Bergey, & Lynda Aiman-Smith, "Forced Distribution Rating Systems And The Improvement Of Workforce Potential: A Baseline Simulation," *Personnel Psychology*, 58 (2005).
70. Olson & Davis (2003); Grote (2005).
71. Scullen et al. (2005).
72. Grote (2005).
73. Grote (1996).
74. G. Pauler, M. Trivedi, & D.K. Gauri, "Assessing Store Performance Models," *European Journal of Operational Research*, 197 (2009), 249–259.
75. See, e.g., Reichheld & Rogers (2005).
76. Pauler et al. (2009).
77. M. Beer & M.D. Cannon, "Promise and Peril in Implementing Pay-For-Performance," *Human Resource Management*, 43, 1 (Spring 2004), 3–48; B. Dowling & R. Richardson, "Evaluating Performance-Related Pay for Managers in the National Health Service," *The International Journal of Human Resource Management*, 8, 3 (June 1997); D. Eskew, R.L. Heneman, & M. Fisher, "A Survey of Merit Pay Plan Effectiveness: End of the Line for Merit Pay or Hope for Improvement?" *Human Resources Planning*, 19, 2 (1996), 12–19.
78. Shah & Murphy (1995); Gallagher (2008); T. Murphy and J. Margulies, "Performance Appraisals," Paper presented at ABA Employment and Law Section Conference, Equal

- Opportunity Employment Committee, Needam Heights, MA (March 24–27, 2004); Grote (2005); Anderson & Cochenour (2001); Norton & Kaplan (1996); Eccles (1991).
79. Gallagher (2008); Murphy & Margulies (2004); Grote (2005); Anderson & Cochenour (2001); Norton & Kaplan (1996); Eccles (1991); Winsor (1984); Thomas & Bretz (1994); Bretz et al. (1992).
80. Shah & Murphy (1995); Bretz et al. (1992); Longnecker (1987).
81. Ami B. Curtis, Richard D. Harvey, & Daran Ravden, “Sources of Political Distortions in Performance Appraisals: Appraisal Purpose and Rater Accountability,” *Group and Organization Management*, 30 (2005), 42–60; Longnecker (1987).
82. Tayla Bauer & Berrin Erdogan, *Organizational Behavior* (Irvington, NY: Flat World Knowledge, 2009); Curtis et al. (2005); Fried et al. (1999); Bretz et al. (1992); Grote (1996).
83. Bretz et al. (1992)
84. Arvey & Murphy (1998); Williams (1992); H.J. Bernardin, “Increasing the Accuracy of Performance Measurement: A Proposed Solution to Erroneous Attributions,” *Human Resource Planning*, 12, 3 (1989), 239–250.
85. Thomas & Bretz (1994).
86. Shah and Murphy (1995).
87. Gallagher (2008); A.J. Rucci & S. P. Kirn. “The employee-customer-profit chain at Sears,” *Harvard Business Review*, 76, 1 (1998), 82–97; Thomas & Bretz (1994); Bretz et al. (1992); S. Gebelin, “Employee Development: Multi-rater feedback goes strategic,” *HRFocus* (January 1996).
88. Flannery et al. (1996); Gallagher (2008); D.L. Bohl, “Minisurvey: 360-Degree Appraisals Yield Superior Results, Survey Shows,” *Compensation & Benefits Review*, 28, 16 (1996), 16–19.
89. Dennis P. Bozeman, “Interrater Agreement in a Multi-Source Performance Appraisal: A Commentary,” *Journal of Organizational Behavior*, 18 (1997), 313–316; Grote (1996); Flannery et al. (1996).
90. Grote (1996).
91. Bozeman (1997); Grote (1996); Manuel London & James W. Smither, “Can multi-source feedback change self-evaluation, skill development and performance?” *Personnel Psychology*, 48 (1995), 375–390; John F. Milliman, Robert A. Zawacki, Carol Normal, Lynda Powell & Jay Kirksey, “Companies evaluate employees from all perspectives,” *Personnel Journal*, (November 1994), 99–103.
92. Bozeman (1997); Grote (1996); Jane A. Layman, “360 Degree Feedback Can Turn on You,” *Fortune*, (November 1994); M. De Lange, I. Fourie & I.J. Van Vuuren, “Reliability of Competency-Based, Multi-Dimensional, Multiple-Rater Performance Ratings,” *Journal of Industrial Psychology*, 29, 2 (2003), 39–48.
93. David Antonioni, “Designing an effective 360-degree appraisal feedback process,” *Organizational Dynamics*, (April 1996), 24–38; Maxine Dalton, “Multi-rater feedback and conditions for change,” *Consulting Psychology Journal: Practice and Research*, 48,1 (1996), 12–16.
94. Thomas & Bretz (1994).
95. J. Welch & S. Welch, “The Importance of Being There,” *Business Week*, 4030 (April 2007), 92.
96. Barbara Crystal, “The 360 Degree Assessment,” *Healthcare Executive*, 9,6 (1994), 18–21; Matthew Budman & Berkeley Rice, “The Rating Game,” *Across the Board* 31,2 (1994), 35–38; Mary N. Vinson, “The pros and cons of 360 degree feedback: making it work,” *Training and Development* (1996) 11–12.
97. J.L. Fahr & J.D. Werbel, “Effects of Purpose of the Appraisal and Expectation of Validation on Self-Appraisal Leniency,” *Journal of Applied Psychology*, 71 (1986), 527–529.
98. Bretz et al. (1992); Bretz et al. (1994).
99. PR Newswire, “Formal Evaluations Improve Employee Performance,” (May 24, 2010). Cross-industry surveys from the 1990s suggest similar patterns. See, e.g., Smith et al. (1996); Shah and Murphy (1995).
100. Thomas & Bretz (1994).
101. Shah & Murphy (1995).
102. Keith T. Jones & Jay S. Rich, “Appraising Employee Performance Evaluation Systems,” *CPA Journal* (February 2008).
103. Welch & Welch (2007).
104. S. Godbehere, “Measuring Staff Performance,” *Business Credit* (November 2005).
105. Ellen M. Heffes, “Aligning Employee Performance with Organizational Goals,” *Financial Executive*, 25, 9 (2009).
106. Welch & Welch (2007); Grote (1996).
107. Thomas & Bretz (1994).
108. F. Nikols, “Performance Appraisal: Weighed and Found Wanting in the Balance,” *Journal for Quality and Participation*, 30, 1 (2007), 13–16.
109. Nikols (2007); Grote (1996); Winsor (1984); Nohria (2008); Murphy & Margulies (2004); Thomas & Bretz (1994).
110. Thomas & Bretz (1994).
111. A. Field, “How Benchmarks, Best Practices, and Incentives Energized PSE&G’s Culture and Performance,” *Balanced Scorecard Report: Case File* (Harvard Business School Publishing, 2008).

112. Nikols (2007); B.G. Mani, "Performance Appraisal Systems, Productivity, and Motivation: A Case Study," *Public Personnel Management*, 31, 2 (2002)141-159; Thomas & Bretz (1994); Winsor (1984).
113. Winsor (1984), Roberts & Reed (1996).
114. Grote (1996).
115. Winsor (1984).
116. Cleveland et al. (1989).
117. Thomas & Bretz (1994).
118. Dávila & Simons (1999)
119. National Performance Review (1997).
120. Grote (2005); Olson & Davis (2003); Thomas & Bretz (1994).
121. Lowery et al. (2002).
122. Bryan C. Hassel and Emily A. Hassel, *Improving Teaching through Pay for Contribution*, (Washington, DC: National Governors Association Center for Best Practices, 2007).
123. Wendy R. Boswell and John W. Boudreau, "Separating the Developmental and Evaluative Performance Appraisal Uses," *Journal of Business and Psychology*, 16, 3 (2002), 391-412; Bretz et al. (1992).
124. Grote (1996).
125. Nikols (2007); Thomas & Bretz (1994); Winsor (1984).
126. Grote (1996).
127. Everatt et al. (2004).
128. Spencer et al. (1997).
129. Zenger (1992); Banker et al. (1996); Bouwens & van Lent (2003); Beer & Cannon (2004); Lowery et al. (2002); Hay Group (2004).
130. Dulewicz & Young (2009).

