SAP SuccessFactors

SAP White Paper
Performance Management

Total Workforce Performance Management

Using talent calibration to effectively manage the reality that all employees are valuable but some employees are more valuable than others

Table of Contents

Lauren Pytel and Steven Hunt

This paper discusses research studying methods for total workforce performance management. The term "total workforce" is used to emphasize performance management processes designed to guide decisions regarding the management of groups of employees. This can be contrasted to methods designed to manage and develop the performance of individual employees in isolation. While aspects of total workforce and individual performance management overlap, certain elements of total workforce management are inherently distinct from methods used for individual performance management. Foremost is the need to manage individual differences in performance found across different employees working in the same group. That is the focus of this paper.

How to fairly and accurately assess employees using calibration

All employees do not perform at the same level[1]. Every employee contributes different levels of value to the organization due to differences in productivity, skills, potential, or any number of other job-relevant factors[2]. These differences are not minor. Contributions made by high performing employees can be several times greater than contributions made by solid or "average" employees[3]. And companies that manage, develop and invest in employees considering their relative contributions significantly outperform companies that treat employees as though they all provide equal value[4]. Recognizing that some employees are more valuable than others isn't just good for companies; it is also good for employees. The use of consistent, transparent methods to assess and reward employee contributions is a key factor affecting employees' perceptions of justice, fairness, and equity[5].

Managing differences in employee contributions is critical to maximizing company performance. But accurately assessing differences in employee contributions is one of the most sensitive and difficult areas of human capital management[6, 7]. People know not everyone contributes equally. But people don't always agree on how their contributions should be evaluated. Many people also find it uncomfortable to have their performance compared against their peers. Employees, particularly those at the "lower end" of the performance distribution, may experience considerable stress from an assessment process that compares their contributions with their coworkers[8]. But it is possible to develop an effective process for assessing employee contributions. Research on employee justice has shown that most employees, even those who may be struggling in their roles, can accept assessment results as being fair provided they understand how they were assessed, believe the process was accurate and consistently applied, and that the results were delivered in a sensitive manner[5]. The challenge is this requires companies to have an accurate and fair method to assess employee contributions—something far easier said than done.

Group-based assessment methods, referred to as "calibration sessions", provide a promising avenue for assessing and managing differences in employee contributions. We define calibration sessions broadly as "meetings where organizational stakeholder discuss the relative contributions of employees." Calibration does not imply

forced ranking or any other specific assessment technique. Despite the growing popularity of calibration, little research has studied the effectiveness of different calibration methods. This paper addresses this issue with a mixture of qualitative research conducted with SAP SuccessFactors customers that use our solutions for different types of calibration, and psychological research of performance appraisals and group decision-making. We provide evidence for the advantages of calibration, discuss critical design questions for creating calibration processes, and offer suggestions on using calibration to accurately and fairly assess employees. We also provide examples showing how technology is enabling companies to implement calibration in more effective ways. This paper is not about technology, but many methods discussed in this paper could not be implemented in a scalable and sustainable fashion without modern, cloud-based technology solutions.

The paper is divided into three sections:

1. **Methods for assessing employee contributions.** An overview of the four most common methods used to assess employee contributions, and a discussion of why calibration sessions are often the most effective of the four.
2. **Defining calibration.** A review of what calibration involves and what makes it an effective method for assessing differences in employee contributions.
3. **Calibration design.** A discussion of critical design questions that must be addressed to build an effective calibration process that reflects the specific needs of your organization.

The goal of this paper is to provide information companies can use to implement group based calibration processes that fairly and effectively address the reality that all employees are valuable but some are more valuable than others.

Part I. Methods for Assessing Employee Contributions

All companies use some method of assessing employee contributions . As long as company leaders want to know "who are the high performers", companies will evaluate the relative contributions of employees. However, the methods used to assess employee contributions vary widely from one organization to the next. Some companies have no formal method to assess performance. These companies rely on unstructured, intuitive opinions and hushed conversations about which employees are more valuable than others. We do not recommend the use of such poorly defined methods since they are likely to be less accurate, less fair, and less useful than more standardized methods. Granted, unstructured performance assessment methods might be sufficient for very small companies where everyone knows everyone else (e.g., fewer than 50 employees). But they are unlikely to work in larger companies where there are considerable legal and financial risks associated with having no defined method to measure employee performance[10]. Virtually all larger companies have some established processes to assess employee contributions . These processes typically use some combination of the following four methods to define and measure the value employees provide to the organization:

1. Contractually defined criteria. This involves defining employee value based on criteria such as job tenure, certification testing, or educational degrees. This method is widely used in unionized workforces, but is not considered an effective method for accurately measuring true employee performance. For example, the ending of the "honeymoon phase" associated with longer tenure tends lead to a decrease in organizational commitment and performance levels[11, 12, 13, 14].

2. Objectively measured criteria. This method works for jobs where there are direct links between individual employee actions and clearly and easily measurable work outcomes. Examples include jobs that measure performance based on individual sales revenue, piece rate productivity, or absence of product defects. A challenge of using objective performance criteria is that this form of performance criteria is not available or particularly relevant for many jobs. Most objective performance criteria are also influenced by factors that are outside the control of individual employees. For example, how much sales revenue an employee generates depends largely on the sales territory they are assigned and the number and type of customers it contains.

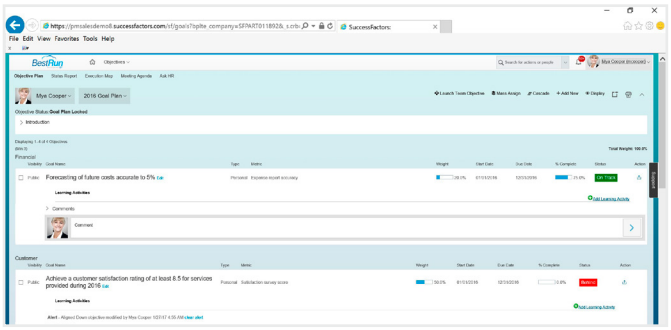


Figure 1: Technology illustration for objective measured performance criteria. This screen illustrates how objective criteria can be incorporated into goal management technology. In this example an employee's performance level is automatically calculated based on attainment of specific financial measures.

3. Individual manager evaluations. This typically involves managers rating employee performance based on an annual, quarterly or monthly schedule. Historically, this has been one of the most common methods used to measure employee contributions. But this method suffers from significant limitations. First, managers may be affected by biases that impact the accuracy of their evaluations. Second, managers have a limited perspective into the contributions of their employees[15]. That is, managers can only evaluate employees based on their view of the employee's performance and this view may not accurately represent the full nature of the employee's contributions. Third, the act of collecting manager ratings is often viewed as a highly burdensome and administrative task.

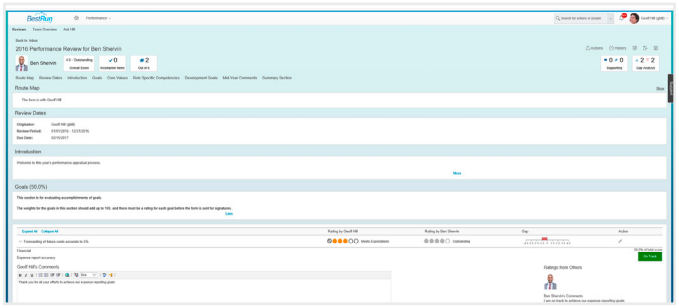


Figure 2: Technology illustration for individual manager evaluations. This screen shot illustrates one of the many ways that manager ratings can be collected through the use of performance management technology.

4. Group calibration sessions. This method involves groups of organizational stakeholders collectively meeting to discuss and evaluate employee contributions. These sessions are sometimes called "talent review meetings". We believe that the use of calibration sessions, when done appropriately, is generally superior to the prior three methods of performance assessment. One reason is because data from contractually defined criteria, objectively measured criteria, and individual manager evaluations can all be incorporated into calibration sessions.

There are several additional reasons why calibration sessions are likely to be more fair, accurate and efficient than other assessment methods:

- **Fair:** When calibration processes are transparently communicated, employees are less likely to feel as though their fate is in the hands of a single manager who may or may not have an accurate perception of their value. Including multiple stakeholders in the assessment helps ensure employees' full range of contributions across the company is being accurately and fairly considered.
- **Accurate:** According to group decision-making research, groups can act as a "check & balance" system against individual biases. Groups with members of diverse backgrounds and perspectives have also been shown to make higher quality decisions than individuals[16].
- **Efficient:** If done right, calibration need not be a drain on organizational time and resources. Calibration sessions can provide leadership with extremely valuable conversations about business and talent within the organization in a relatively short amount of time.

We believe calibration tends to be more effective than other methods of performance assessment for most jobs. But calibration does not, and as we will argue, should not look the same across every organization or every job. There are different methods of conducting calibration, different purposes it may serve, and other organizational factors that may affect its value and usefulness. The purpose of this paper is not to reveal the one and only way calibration must be performed, but to support an organization's use of calibration in a thoughtful and strategic way.

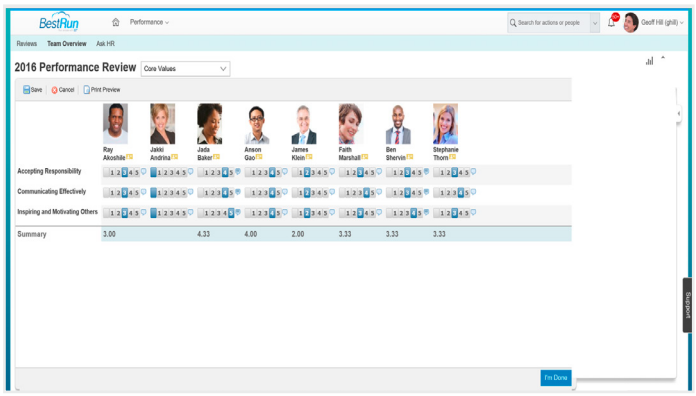


Figure 3: Technology illustration for calibration talent review. This is one of several screen shots provided that show how calibration methods can be supported through technology. In this example, a manager or group of managers are able to simultaneously review and compare the performance of multiple employees on a range of different performance dimensions and job relevant characteristics.

Part II. Defining Calibration

According to Merriam Webster's dictionary, the word "calibrate" means to measure against a standard. In the context of talent management, "calibration" refers to methods used to ensure that evaluations of employee contributions are based on a common and appropriate set of standards. We use the word "contribution" intentionally because calibration is used in a variety of ways. Perhaps the most common use of calibration is to assess differences in employee performance in their current role. Calibration is also frequently used to assess differences in employee potential to move into future roles. It can also be used to assess the criticality of an individual employee's skills and capabilities to the operations of the company. For example, determining if certain employees possess crucial technical knowledge that could not be easily replaced if they left the company. Calibration can also be used to determine how to allocate compensation, job opportunities, or development resources across employees in ways that will maximize the return on investment for the organization.

Calibration creates an "equal playing field" by providing managers with reference points, common standards and shared criteria against which to make judgments about employees. Rather than relying on the subjective opinions and attitudes of a single manager, calibration brings in the opinions of multiple people with different viewpoints and perspectives to ensure that employees are evaluated based on their actual behaviors, skills and accomplishments. This is particularly important for reducing idiosyncratic biases found across managers and reducing subconscious biases that can negatively impact women, ethnic minorities and other historically under-represented groups[17, 18].

Calibration also addresses one of the most common problems associated with assessments made by individual managers: the tendency of managers to categorize all employees as being roughly the same. Left on their own, many managers seek to avoid confrontation, hurt feelings, etc. by placing most or all their employees at the same general level. They may also categorize the majority of employees in a way that implies their contributions are all "above average", even though by definition this is not possible. A fundamental part of effective calibration is using group-based decision making methods to challenge and help managers differentiate between outstanding, solid and below average contributors. Some common methods for doing this include:

- **Clear assessment criteria.** Assessing employees using behavioral definitions of job performance, metrics tied to specific, measurable job goals, and/or well defined attributes describing key skills and capabilities. During calibration sessions managers are challenged to describe and categorize employees relative to these criteria.
- **Expected rating distributions.** Some calibration sessions encourage managers to distribute employee ratings to fit a pre-defined distribution (i.e., 15% "fails to meet expectations", 35% "meets expectations", 35% "exceeds expectations", and 15% "greatly exceeds expectations"). The use of recommended distributions can be an effective means to encourage managers to critically evaluate employees' relative contributions. But strictly requiring managers to "force rank" employees into different categories can be highly problematic. Force ranking may initially be effective at increasing workforce productivity if a company has a high percentage of under-performing employees that it has not addressed. But its value wears off quickly as the company begins to manage out poor performers[19, 20] and over time it can significantly damage workforce quality and employee morale[19, 20, 21]. Consequently, we do not recommend using strict forced ranking unless there is a very specific situation that requires it (e.g., financial restructuring that necessitates significant reductions in workforce cost). Even then, it should be used very carefully.
- **Reviewing and discussing categorization ratings.** One of the hallmarks of effective calibration sessions is the use of discussion and dialogue to clarify what employee behaviors, skills and accomplishments are critical to the success of the organization. Research has shown that when raters know they will be asked to justify assessment ratings, they put more care into consideration of performance behaviors when making their assessments[22]. The result is more thorough, consistent, unbiased assessment of employees. The most common type of rating review involves an employee being rated by his or her manager, and then that manager's manager reviewing the rating to ensure it aligns with the company's performance definitions and/or rating distribution guidelines. This method is a form of calibration, but it is prone to allowing bias to affect ratings since both managers may share the same perspectives[23, 24]. Calibration sessions reduce this risk of bias by bringing together a larger group of individuals composed of managers, senior leaders, and/or talent management specialists from across multiple teams, departments, or organizations to discuss and debate assessment ratings.

Calibration sessions can leverage all three of these methods to increase assessment accuracy. For example, performance data and ratings from individual managers based on clear performance criteria may be collected prior to a calibration review session. During the session managers are required to explain and discuss their ratings decisions to their peers. In some cases, the discussion may lead managers to change their ratings. Managers may also be challenged if their ratings do not conform to an expected rating distribution (e.g., if a manager rated all their employees as above average they would need to explain to their peers why their team is so strong). This ensures that the final assessments are both accurate and consistent with the views of and observations of other members of the organization.

One of the historical challenges to conducting calibration was the work required to assemble the data on individual employees needed to hold effective calibration sessions. Because of the effort involved, calibration was often only done for a very small number of jobs in a company such as top leadership roles. Or it was done in an overly simplistic manner that emphasized compliance over quality (e.g., forcing managers to comply with strict forced ranking guidelines). Advances in human capital management technology have removed many of the operational constraints that historically limited companies from making much greater use of calibration. As a result, the use of calibration is becoming more common across a much wider range of jobs than in the past.

Given its multiple benefits, we believe that calibration should be used by virtually every organization in some manner (an exception might be made for very small companies). This does not mean calibration should be used in the same way or for the same purpose in every organization. Nor should it necessarily be used for every type of job. But given the value calibration has for ensuring fair, accurate assessments of employee contributions, it is hard to imagine a company that could not benefit from its use in some form.

"We use a 'good cop bad cop' type strategy to push supervisors to really flush out their thought process and justify their ratings further."

"Disagreements are generally where leaders must really justify their view of a person's performance. If I've interacted with the person myself, I might say to the manager, 'well this was my interaction with so and so, and this is what occurred... is this an anomaly? Is this typical behavior for so and so? Things like that."

Part III. Calibration Design

There is no "one best way" to conduct calibration sessions. Methods that work in one company might not make sense for another organization. The following are critical design questions that should be considered when designing a calibration process or conducting calibration sessions for your company:

1. What are you calibrating?
2. What jobs and employees are being calibrated?
3. How much time will be spent on calibration?
4. How will calibration data be used?
5. Which employees will be assessed in different calibration sessions?
6. What methods will be used to structure calibration conversations and decisions?
7. Who will participate in calibration sessions and what are their roles?
8. How will the results of calibration sessions be communicated?

Each of these questions will be discussed in more detail.

1. WHAT ARE YOU CALIBRATING?

It is important to ask "what purpose will calibration serve in my organization?", and to make sure that managers clearly understand this purpose. Calibration may be used for several purposes, including:

- **Managing differences in job performance.** The success of a company ultimately comes down to the performance of its employees. And the best predictor of future performance is past performance. While job performance is critical to business success, it is also notoriously difficult to define and measure. Performance can be defined in terms of past accomplishments, quantitative results, knowledge and skill levels, or any number of behaviors. Managers vary in how they evaluate employee performance, have difficulty differentiating between high and low performers, and struggle to effectively deal with employee performance differences. Left to their own devices, many managers choose to treat all employees as though they perform at the same level, rather than addressing low performers or supporting truly high performers.

"We saw ratings continue to become higher and higher. It's easier to say 'your performance has improved' than the opposite. So, we used round tables to get calibrated ratings back."

"Looking at how managers were rating employee performance, we knew we had to do something. Previously, everyone was getting a '5'. But we knew from corrective actions throughout the year, that these ratings couldn't possibly be accurate. Calibration seems to be working. Rarely do we find employees in the lower two steps for more than a year. They either exit the organization, or more likely, have improved their skills and are meeting expectations."

Calibration addresses these problems by using common definitions and processes for assessing performance, and by fostering conversations that provide guidance and expectations about how performance differences are to be managed. By improving the accuracy of performance ratings, calibration provides the company with valuable data for making talent decisions related to staffing and compensation, tracking metrics such as "turnover of high vs.

low performers", or using performance criteria to evaluate and refine hiring and training methods.

- **Assessing differences in employee potential.** To maintain organizational performance over time, companies must effectively support and invest in the potential of their employees. Accurately measuring employee potential is critical to guiding development decisions related to training, job assignments, or retention incentives. Like performance, the concept of potential can be difficult to define. It depends on a range of factors including past performance, motivation, aptitude, qualifications, and others. Calibration helps ensure that assessments of potential are based on a common, organizational definition and not simply the untested judgment or "intuition" of individual managers. Calibration sessions can also increase visibility of high potential employees and identify opportunities to provide them with job assignments or resources that will help them realize their potential.

  "The problem was that managers thought of 'high potential' in different ways across our organization. There are so many definitions of potential, so the definition would change whether speaking about it in the investment business or the distribution channel, for example. In any case, we weren't quantifying potential consistently across the organization, which is what calibration helped us to do."

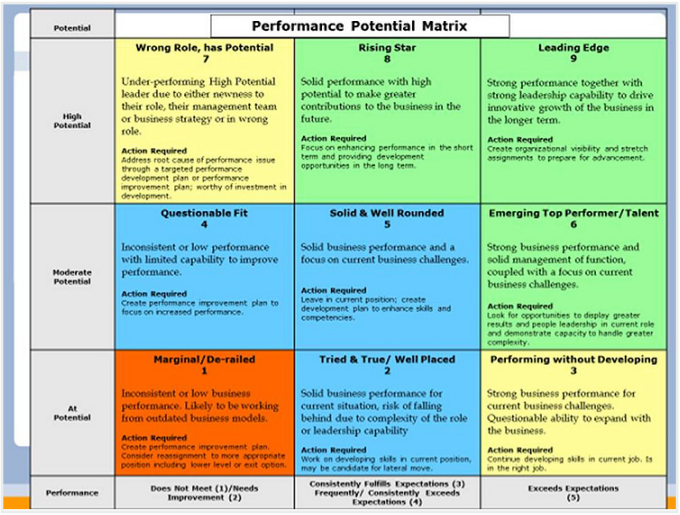
- **Allocating compensation and rewards.** It is appropriate to invest resources into employees based in part on the relative value they provide to the company[26]. Rewarding performance, either through merit increases, bonuses, long-term incentive, or spot awards, has been shown to significantly increase workforce performance over time[27]. But these 'pay for performance' techniques only work if the organization has methods to appropriately determine the level of rewards to be given to different employees. Calibration can be a key part of these methods.

  It is important to stress that using calibration for compensation is different than calibrating performance or potential. While performance and potential are important factors that influence compensation, they are not the only things that affect compensation decisions. For example, current pay levels compared to the market, comparing pay levels between employees in similar roles, previous pay increases, and perceived retention risk are all reasonable factors to consider during a calibration session focused on compensation. None of these factors would make sense to include in a calibration session focused on performance.

  "There are a lot of good employees. But there's only so much money to give out."

It is common and sensible for companies to have multiple calibration sessions with different purposes. For example, a company might hold calibration sessions focused on performance in the first quarter of the year and then hold sessions focused on compensation the second quarter. The compensation calibration sessions might leverage data from the performance calibration sessions, but would also use other data to guide pay decisions. In the third quarter the company might hold another series of calibration sessions focused on assessing employee potential. These sessions might be limited to

employees in certain roles or employee who have achieved some minimum level of performance.



Potential	Performance Potential Matrix		
High Potential	**Wrong Role, has Potential** 7 Under-performing High Potential leader due to either newness to their role, their management team or business strategy or in wrong role. **Action Required** Address root cause of performance issue through a targeted performance development plan or performance improvement plan; worthy of investment in development.	**Rising Star** 8 Solid performance with high potential to make greater contributions to the business in the future. **Action Required** Focus on enhancing performance in the short term and providing development opportunities in the long term.	**Leading Edge** 9 Strong performance together with strong leadership capability to drive innovative growth of the business in the longer term. **Action Required** Create organizational visibility and stretch assignments to prepare for advancement.
Moderate Potential	**Questionable Fit** 4 Inconsistent or low performance with limited capability to improve performance. **Action Required** Create performance improvement plan to focus on increased performance.	**Solid & Well Rounded** 5 Solid business performance and a focus on current business challenges. **Action Required** Leave in current position; create development plan to enhance skills and competencies.	**Emerging Top Performer/Talent** 6 Strong business performance and solid management of function, coupled with a focus on current business challenges. **Action Required** Look for opportunities to display greater results and people leadership in current role and demonstrate capacity to handle greater complexity.
At Potential	**Marginal/De-railed** 1 Inconsistent or low business performance. Likely to be working from outdated business models. **Action Required** Create performance improvement plan. Consider reassignment to more appropriate position including lower level or exit option.	**Tried & True/ Well Placed** 2 Solid business performance for current situation, risk of falling behind due to complexity of the role or leadership capability **Action Required** Work on developing skills in current position, may be candidate for lateral move.	**Performing without Developing** 3 Strong business performance for current business challenges. Questionable ability to expand with the business. **Action Required** Continue developing skills in current job. Is in the right job.
Performance	Does Not Meet (1)/Needs Improvement (2)	Consistently Fulfills Expectations (3) Frequently/ Consistently Exceeds Expectations (4)	Exceeds Expectations (5)

Figure 4: This screen illustrates one of the more common methods used to calibrate employees for the purpose of succession management. Frequently called the "9 box", it assesses employees based on performance in their current role and their future potential to move into roles with greater levels of responsibility. This method is called a 9 box because most companies assess employees on 3 levels of performance and 3 levels of potential. Variations of the 9 box include 25 boxes, 6 boxes, 4 boxes or any other combination one might imagine based on the number of performance and potential categories used. As we discuss later, the 9 box method while widely used and beneficial in many respects, also has some significant limitations.

In addition to using calibration for performance, potential, and compensation, companies are increasingly using calibration to guide operational staffing decisions. For example, following announcement of an acquisition a company might hold a calibration session to determine who to assign to the post-merger integration team. During this session leaders meet to consider which employees to assign to the integration team based on their past experiences, skills and future potential. Technology has made these operational, "spur of the moment" operational staffing calibration sessions much easier to conduct by allowing companies to quickly identify and create pools of employees and to readily review employee performance and potential data.

There are several reasons why it is better to conduct separate calibration sessions rather than trying to calibrate performance, compensation and potential all in a single session. First, holding multiple sessions simplifies calibration by focusing on one type of decision at a time. This tends to improve the results of calibration sessions. Research has shown that the broader the number of topics and individuals included in group decisions, the greater difficulty they may encounter in terms of coordination and decision making[28]. In other words, as more people become part of a conversation and more topics are covered, the greater the risk of confusion and misunderstanding.

Having multiple calibration sessions also allows a company to design each session tailored to its specific purpose. This may include changing the composition of the calibration group, the criteria used for calibration, and the timing and steps used in the overall calibration process. For example, it may make sense to have one group of individuals calibrate past employee performance, but another group to discuss future employee potential. The design and agenda of the session will also change

depending on whether the discussion is about performance, potential, or compensation. And it may make sense to calibrate performance across a broad range of jobs and employees, but only calibrate potential or compensation for a smaller number of people. In sum, it is better to have several types of calibration sessions each focused on a single topic rather than one long session that tries to cover multiple things at once.

2. WHAT JOBS AND EMPLOYEES ARE BEING CALIBRATED?

It usually does not make sense to include all employees or all jobs in every type of calibration session. The following are factors to consider when evaluating which jobs should be included in different types of calibration processes:

- **Impact of performance differences.** Calibration is most valuable when applied to jobs where differences in employee performance have a significant impact on business outcomes. This is true for most jobs. However, there may be some frontline, entry-level jobs where performance is highly routine, to the point that there is relatively little difference between a high performer and an average performer. In these cases, calibration may not make sense.
- **Length of service.** The use of calibration helps organizations to steadily improve the overall quality of the workforce. This is particularly important for jobs where employees are likely to remain with the company for multiple years. Calibration may not make sense in jobs where employment is seasonal or where the average length of service is only 12 months or less.
- **Work contracts.** Certain types of calibration may not make sense or even be legal for jobs where workforce decisions are governed by collective bargaining contracts or other pre-defined qualifications criteria. For example, it would be meaningless to hold calibration compensation sessions for a job where compensation levels are contractually determined based on tenure. But it may still make sense to calibrate performance and potential for these sorts of jobs to identify and support high performing employees, surface and address potential performance concerns, and develop employees who have the aptitude and interest to move into more senior or leadership positions.
- **Career paths.** In many companies, certain jobs provide internal talent pools for the organization. For example, many retail companies leverage frontline hourly staff to identify and develop future store managers. And many manufacturing companies draw supervisors from frontline employees working under union contracts. Calibration sessions focused on potential can add value for these jobs even if performance differences in these jobs are minimal, employees in these jobs tend to have a relatively short average length of service, or pay is determined by contractual defined criteria.

  "We want managers to think of employees as constantly developing no matter where the employee is in the organization. Even if an employee is unlikely to move into a new position, we still want to be thinking about "what's their next best job"?

- **Employment Status.** This is not so much about what jobs to include, but whether to include people who may not have "jobs" in a technical sense. Most companies limit calibration sessions to individuals that are currently employed by the company. But some organizations include individuals that recently quit or were asked to leave

the company, particularly when calibrating on performance. This enables the company to track whether low performing employees are being effectively managed out and whether high performing employees are voluntarily leaving. Companies may also want to include contract employees in some calibration sessions, although there may be legal restrictions that limit the involvement of contractors in activities associated with full-time employment.

Calibration sessions can require a considerable time investment. The value of this investment will vary depending on the nature of different jobs. Table 2 provides a summary of jobs where certain types of calibration may not make sense. For example, calibration methods that make sense for highly skilled, longer-tenured professional positions may not make sense for lower skilled, short-tenured, entry level jobs. This does not mean that calibration never makes sense for certain kinds of jobs, but that the value associated with different calibration methods can change depending on job type.

3. HOW MUCH TIME WILL BE SPENT ON CALIBRATION?

One of the primary ways calibration provides value is through getting managers and other key members of the organization to discuss important topics and decisions related to workforce performance, potential and compensation. The time required for effective calibration discussions is one of its biggest limitations. Calibration sessions do not work if they are overly rushed. But holding numerous lengthy calibration sessions is rarely possible given most companies operational constraints. This is particularly true in large organizations where the sheer number of employees and job types can pose a challenge to running an efficient calibration process.

"It takes at least 30 minutes to hold a good calibration session and usually more. The goal is not just to place people into boxes. It is to spend time talking about why they are in different boxes."

Table 2. Jobs Where Calibration Methods May Have Less Value

	Likely value of calibration methods		
	Performance	Potential	Compensation
Jobs where performance differences are small	Low	Moderate	Low
Seasonal or high turnover jobs	Low	Moderate	Low
Jobs where compensation is set by contracts	Moderate	Moderate/ High	None
Jobs with little internal turnover to new positions	High	Low	High
Contract or former employees	High	None	Low

The SAP SuccessFactors customers we interviewed for this study reported assessing anywhere from 1,000 to over 145,000 employees during their calibration processes. Smaller organizations tended to spend about 6-8 weeks completing their calibration process. Larger, more complex organizations reported spending up to 9 months of the year preparing for and conducting the formal sessions. The length of individual calibration sessions also varies substantially across our customers. Sessions focused on lower level jobs or that involve fewer than 10 employees may be as short as 30 minutes. In contrast, sessions looking at larger numbers of employees or at high impact professional and leadership jobs tend to be 2 to 4 hours in length. Calibration sessions that are focused on a thorough assessment of top leadership and professional employees, or that are examining the total talent strength of large workforces may take up to a day or more.

A critical part of calibration process design is thinking through the time required to conduct a typical calibration session considering how many sessions will need to be conducted across the company. A calibration process that requires more time and resources than an organization can realistically devote is bound to fail. The following variables should be considered when thinking about the time needed to complete a calibration process:

- **How many types of calibration sessions will you conduct?** A tempting way to reduce the time spent on calibration is to combine different types of calibration into a single session rather than holding multiple sessions. For example, calibrating performance, potential and compensation all in the same session. There are two reasons why combining calibration sessions is not wise. First, combined sessions create confusion. People struggle to differentiate between performance, potential and compensation which results in less effective calibration decisions. This undermines the core purpose of calibration which is to improve the accuracy of employee assessments. Second, it may actually take longer. Group decision-making is faster and more accurate when the group is focused on clearly defined criteria and outcomes[29]. We believe it is usually better to conduct a single type of calibration well than risk doing multiple types of calibration poorly. It also need not be an "all or nothing" proposition. For example, a company might conduct calibration sessions for performance on all employees, but limit calibration sessions for potential to those employees who are above a certain level of performance.
- **How many employees will be assessed in each calibration session?** Increasing the number of employees assessed in each session reduces the total number of sessions needed. However, adding employees can increase the length of each individual session and at some point, will lead to "decision fatigue". Decision fatigue occurs when groups are making the same sorts of decisions over and over and significantly reduces the accuracy of group decisions[30, 31]. The quality of the calibration sessions can also suffer as a result of combining together groups of employees that should be calibrated separately. In sum, a host of factors should be considered when deciding how to combine groups of employees. Later in the paper we discuss in more detail some of the factors to consider when determining the number and types of employees to include in a calibration session.

- **How much time will be spent discussing individual employees in each calibration session?** A major purpose of calibration is to encourage conversation about employee contributions. At the same time, there is a point of diminishing return when it comes to talking about an individual employee. How much time is enough or too much is likely to depend on the nature of the employee and their role in the organization. For calibration sessions focused on senior leadership roles, it is not uncommon to spend 30 minutes talking about a single employee. But for most roles, spending more than 15 minutes on one employee would probably be considered excessive. And less than 3 minutes may be adequate for many employees. However, it is critical that adequate time be scheduled for calibration sessions to avoid the risk of suppressing valuable conversation merely to stick to a pre-defined schedule.
- **Will you have dedicated calibration session facilitators?** As we will discuss later in this paper, there are many reasons why it may make sense to have a dedicated session facilitator. Skilled session facilitators are critical to keeping calibration conversations focused and efficient while remaining rigorous and effective. Research has shown that a skilled facilitator can increase group decision quality[30] while simultaneously reducing time needed for discussion[32].

Ultimately, the only correct answer to the question "how much time will calibration take?" should be "enough to make accurate and appropriate assessment decisions without overly burdening business operations". Calibration sessions can take as little as 15 minutes or as much as a full day depending on the topic, the number and type of employees, and the structure and facilitation of the session. How long is long enough should be a subject for discussion, keeping in mind that a) few topics have a bigger impact on business performance than workforce performance and potential, and b) good calibration conversation is not just about the capabilities of employees; it is also about what capabilities are needed by the organization. Calibration is not just a conversation about talent; it is also a conversation about the strategic needs of the business. And that is a conversation worth spending some time discussing.

4. HOW WILL CALIBRATION DATA BE USED?

Calibration sessions provide value in two ways. First, the dialogue during calibration sessions provides managers and leaders with a greater understanding of the talent within the organization and relationships between the company's business needs and the capabilities of its workforce. Second, calibration sessions provide data segmenting the workforce into critical categories based on the characteristic and contribution of individual employees. This data is highly valuable for improving organizational talent management decisions, provided it is used as part of the decision-making process!

It is important to communicate how calibration data will be used prior to conducting calibration sessions. Being transparent about the purpose and use of calibration data is critical for building employee-manager trust regarding the fairness and equity of the company's talent management methods[8], [33]. In addition, if managers know that the data from calibration sessions will be used to make important decisions impacting their teams and the organization, they will likely put greater effort into ensuring the data generated through calibration discussions is accurate[34].

There are several ways that calibration data can be leveraged by the organization. The first and most obvious is to use the data to make talent decisions about individual employees. For example, using performance calibration data to guide decisions related to compensation. Or using potential calibration data to guide decisions related to development and staffing. The second way is to use the data to gain insights into the nature of the workforce. For example, determining the percentage of the workforce that is exceeding performance expectations or that is considered ready to take on greater levels of job responsibility. The third way data can be used is to measure workforce strength. For example, measuring employee retention by performance level to determine if high value employees are turning over at a faster rate than lower value employees. Or, assessing if employees in certain demographic groups are disproportionately categorized as being at lower or higher levels of performance and potential. Last, calibration data can be used to evaluate the effectiveness of different managers or talent processes. For example, do certain managers excel at retaining and developing high performing employees. Or do certain recruiting methods lead to hiring higher performing employees.

The more calibration data is used the more effort will go into calibration discussions, and the more accurate and valuable the data will become. The key is to start using the data and to be sure that managers and employees know how the data is being used.

5. WHICH EMPLOYEES WILL BE ASSESSED IN DIFFERENT CALIBRATION SESSIONS?

One of the most important parts of designing a calibration process is determining which employees will be assessed in the same session. Addressing this topic requires managing the risk of putting too many employees in a single session and not having adequate time available to discuss the employees' individual contributions or combining employees who cannot be effectively compared against one another. The following are some crucial questions to think through when addressing this topic.

- **How many employees should be included in a session?** Figuring out how to create calibration groups of manageable size and appropriate composition is critical to making calibration work. For operational reasons alone, it may make sense to set a maximum number of employees for a single calibration session. Many of our customers reported setting a cap on the maximum calibration group size (i.e., 50 or 80 employees maximum), or having a specific time allocated to discuss each employee (e.g., 3 to 5 minutes). Be realistic about the time your organization is willing and able to devote to calibration. If you want to include 200 employees in a single session, be prepared to spend the better part of a day on the discussion or be comfortable knowing that only a small percentage of employees in the session will actually be discussed.

One strategy for calibrating large numbers of employees is to allow the size of calibration sessions to increase over the course of the process. Initial calibration sessions done within specific functions or regions may be limited to smaller numbers of employees to ensure adequate time to discuss every individual. The results of these sessions are then aggregated to create sessions that span multiple functions and regions and include much larger numbers of employees. These larger sessions use a different calibration process where discussion is limited to a subset of employees (e.g. only discussing outstanding performers or high potentials).

"Some companies try to put hundreds upon hundreds of employees into these sessions. But calibration is really intended to allow for discussion about each individual person, or at least close to each person, during a single session."

It is also important to consider the minimum number of people to assess in a single calibration session. Calibration involves comparing employees on different attributes and placing them into categories based on their perceived value or impact on the business. The assumption underlying calibration might be summed up as "not everyone in a group contributes at the same level and there is value in knowing who is contributing the most or least to the organization". Calibration provides the most value when you are considering the performance of at least 3 or more people. You could have a calibration session with just one employee, but it wouldn't be true calibration considering no comparison is being made to other employees. The issue of how many categories should be used and questions about limiting the number of employees who can go into each section when using forced distribution is discussed later.

Some companies set a minimum number of people to include in calibration sessions. This ensures a basis of comparison for the calibration discussion. It also reduces the time spent conducting calibration sessions that only assess small numbers of people. The issue of minimum calibration size becomes particularly important if your company requires that employees included in a calibration session must be placed in different categories (i.e., everyone can't be given the same rating). Research has shown that these sorts of forced distribution methods do not work effectively unless you include at least 30 employees in each calibration session[47]. The issue is one of mathematics. The concept of forced distribution assumes that not everyone performs at the same level. But in small groups of employees, it is possible that everyone may truly be at the same level.

In summary, it is important to have a set of guidelines around setting minimum and maximum numbers of employees to include in calibration sessions. This doesn't mean that all calibration sessions should be equal in size. It is quite common for initial calibration sessions to be conducted with smaller groups of employees representing specific departments or regions, and then to conduct larger calibration sessions that combine these groups together to look at talent across the company. However, the structure of these session should change as the number of employees increases. For example, small group calibration session might should encourage conversation about every employee while large group sessions might focus on only those

employees falling into the upper or lower categories of performance, potential or compensation.

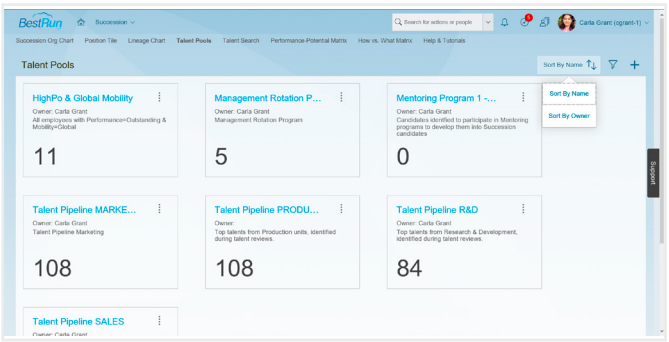


Figure 5: Technology illustration for selecting employees to include in a talent review session. The screenshot shows that individual employees can be selected for inclusion in a calibration talent review based on their organization, performance level, tenure, interest in relocation, and a range of other criteria.

- **Which employees should be calibrated together?** The most common method used to determine which employees to include in a single calibration session is to look at the company organizational chart and combine employees who report up to the same managers or leaders. The advantage of this method is that it is very simple. The problem is it assumes employees who report to the same person on an organizational chart work together or perform similar functions. This assumption is not always true. In matrixed organizations or companies with highly specialized roles, two employees who report to the same manager might perform extremely different tasks and rarely interact with each other. It may be difficult to compare these employees against each other in a calibration session. It is therefore important to consider whether criteria other than reporting structure should be used to guide selection of employees to include together in calibration sessions. The following are criteria companies have used to determine which employees to calibrate together:

- **Reporting structure.** This is probably the most common method used to determine calibration groups. But it need not be limited to formal reporting structures. Companies may also group employees together based on matrixed reporting relationships.
- **Job categories.** Employees can be grouped together based on the types of jobs they perform. For example, calibrating first level managers together as a group or calibrating engineers together as a group. It is also common to group employees based on their functional area. For example, calibrate finance personnel together in one group and marketing personnel together in another group.
  - **Geographic location.** Some companies group employees based on the region or office they work from. This approach makes it easier to conduct in-person calibration meetings since everyone is likely to work in the same place. But it can be problematic because employees in one region may perform very different types of work.
  - **Employee attributes.** Companies may group employees together who share common attributes related to career interests or qualifications. This is particularly common when defining groups when calibrating for potential. For example, conducting calibration sessions for employees who have indicated a willingness to relocate, an interest in advancement, more than one year of

tenure in their current role, and/or performance above a certain minimum standard.

It may make sense to use multiple criteria to guide the selection of employees for calibration sessions. For example, conducting a calibration session looking at the potential of all first level managers in sales working in Europe with at least a year of tenure. Technology can be very valuable for this task because it enables companies to quickly identify and select employees who meet different criteria. On the other hand, if the selection criteria become too complex, it may become operationally difficult to scale calibration across the company in a coherent fashion. Ultimately, it comes down to figuring out what makes the most sense given the nature of the company and the goals you wish to achieve through the calibration sessions.

The decision of how to combine employees into calibration sessions is not a trivial one. Determining how many employees to include in calibration sessions and which employees to calibrate together has a significant impact on how calibration sessions are conducted. Calibration methods that work for small groups may not work for large groups. And it can be difficult to compare employees if they come from radically different jobs. Similarly, the people doing the rating may struggle to have effective discussions about employees who they do not know or who perform roles they do not understand. At the end of the day, there is no "ideal way" to determine the composition of calibration sessions; every approach will have advantages and disadvantages. What is important is to have a clear viewpoint and rationale behind the decision of which employees to calibrate together and to explain this to employees and managers so they understand why the sessions are structured the way they are.

6. WHAT METHODS WILL BE USED TO STRUCTURE CALIBRATION CONVERSATIONS AND DECISIONS?

This question focuses on the methods you will use to guide conversations and decisions during the actual calibration sessions. Calibration involves bringing a group of people together, usually managers and company leaders, to discuss and assess the contributions of other people in the company. Calibration is fundamentally a group decision making process. A lot of research has shown that groups make better decisions when they follow well-defined processes[35]. This involves defining the criteria used for the assessment, defining the order you will use to assess employees, structuring the dialogue used to guide the assessment, and clarifying the expected outcomes of the assessment process. We discuss each of these below.

- **Assessment criteria.** Most calibration sessions focus on categorizing employees based on evaluations of current job performance, estimates of future job potential and/or recommended compensation investments. Calibration discussions tend to be more efficient and effective if the categorization criteria are clearly defined and communicated in advance of the session. Table 3 lists some common criteria used to categorize employees into different levels of performance, potential and compensation. These criteria may not make sense for every organization. But it is recommended that every organization identify, define and communicate these types of criteria in advance of conducting calibration reviews.

- **Order of assessment.** Research has shown that the order in which individuals are assessed influence how they are assessed[36]. For example, research studying the "contrast effect" has shown that when an employee receives a low assessment the next person assessed is more likely to receive a higher assessment[37, 38]. And research examining "rating fatigue" suggests that more time will typically be spent discussing the first person who is assessed compared to the last person. It is probably impossible to assess employees in an order that won't have some impact on the assessment results. But the level of error created by the order of assessment can be reduced using a few simple tips. First, create a standard process that defines the order of assessment. This will ensure consistency across calibration sessions. Second, avoid ordering assessment using arbitrary criteria such as an employee's last name or the manager they

Table 3. Common criteria used to guide calibration categorization decisions

Current Job Performance	Future Job Potential	Recommended Compensation Investment
• Level of achievement of different job goals • Difficulty or importance of job goals achieved • Displaying specific positive or negative job relevant behaviors • Manager evaluation of performance • Recommendations from peers, direct reports, and/or customers • Achievement of major development goals; acquisition of new skills and certifications • Organizational commitment as demonstrated through job tenure and length of service	• Performance in current job with emphasis on specific behaviors or goals associated with success in possible future roles • Employee career interests (e.g., interest in job opportunities, willingness to relocate) • Acquisition of specific job experiences (e.g., managing P&L, international work) • Acquisition of skills, certifications or educational qualifications • Manager evaluation of potentia • Measures of potential gathered via standardized assessment tools or assessment centers	• Current job performance • Future job potential • Current pay relative to external market rates • Current pay relative to internal peers or job level • Previous pay increases • Job tenure • Achievement of specific job goals that had a major impact on company performance • Acquisition or possession of specific skills, knowledge or job qualifications viewed as critical to future business operations • Manager rating of retention risk

happen to report to. Using these criteria could unfairly impact the results of people with certain names or who happen to report to certain managers.

The following are some methods companies have used to determine the order of assessments that can help lessen the risk of biased results.

- **Anchoring.** This is probably the most common method. In this method, people assessing employees in the calibrations session start by writing down privately the employee they feel has the strongest performance, potential or compensation value (depending on the type of calibration session being conducted). One person is selected at random to share the employee they identified and explain why they identified them. The group then discusses the assessment of this employee and places them in an initial category. Another person is then selected to share who they identified and the process continues.

  Anchoring can start at the high end or low end, although most companies start on the high end. It is not recommended to jump back and forth from high to low and back again. This could both create confusion and could heighten bias caused by the "contrast effect" when employees look better or worse based on who was assessed before them. When calibrating large groups of employees, it is quite common to spend more time discussing employees during the early phases of anchoring and then spend less time as the process progress. This is a natural result of people developing a shared sense of criteria for evaluating employees.

- **Job Categories.** This method is often used when calibrating employees who come from multiple jobs, functions or geographies. Employees are grouped together based on some aspect of their work. For example, working in the same job, being part of the same team, or working in the same geographical area. Grouping employees together based on similar work characteristics can make it easier to compare employees against one another since they perform similar tasks or work in the same environment. Although this method does create a risk of bias because some employees may look better or worse based on the level of performance found in their specific group or team[10].
- **Randomization.** In this method, employees are selected at random to be assessed. This method works well for calibration sessions that are assessing relatively small numbers of employees.

When conducting calibration sessions involving very large numbers of employees, it may make sense to limit the discussion to a subset of employees in the group. For example, only discussing employees who meet certain pre-existing criteria such as being identified as high performers in previous calibration sessions. In these situations, you can still apply these ordering methods but only use them for the sub-group of employees that are going to actively discussed. Last, note that these ordering methods can be used in conjunction with each other. For example, you might use job categories to identify the initial group of employees to be assessed and then order those employees for assessment using randomization.

- **Structuring the assessment conversation.** It is useful to develop a script to guide the assessment conversation for each employee. This ensure all employees go through a similar assessment process. For example, when calibrating based on performance you might start the discussion about each employee with a standard set of questions such as: "What outstanding accomplishments or behaviors has this employee achieved in their role? What things does this employee do that make them more effective than their peers? What things make them less effective? What would this employee have to do to reach the next highest level of performance?" Following a structured conversation process is fundamental to consistent, accurate measurement which is one of the primary goals of calibration.

  You may also want to establish some process to determine when to skip, start or stop conversation about employees. For example, always asking the question "does anyone have additional information that they believe might impact how this employee as has been categorized?" before moving on the next employee. This lessens the chance of employees being unfairly overlooked or otherwise receiving inadequate attention during a calibration session.

- **Expected calibration outcomes.** The purpose of calibration is to recognize, accurately assess and understand differences in employee contributions so the company can more effectively and fairly manage its workforce. The fundamental assumption of calibration is that "not all employees contribute equally". A risk when conducting multiple calibration sessions is that different groups of raters may interpret the term "equal" differently. Raters in one calibration session may be comfortable grouping most employees into the same categories as though they are equal, while raters in other sessions may work hard to differentiate employees from each other. The best way to lessen this risk is to define the number of categories used in the assessment and to set expected rating distributions across these categories. For example, setting a goal in a performance calibration session to place employees into 5 different categories such as "record breaking, outstanding, solid, below expectations, serious concerns" and then communicating that the company expects to see about 5% to 10% of the employees in the highest category, 25% to 30% in the next highest category, and so forth.

  As mentioned earlier in the paper, there is a big difference between communicating "expected" rating distributions versus using "forced distributions" that require that a certain percentage of employees go into different categories. The use of forced distribution can create significant problems and should generally be avoided[21]. However, communicating expected rating distributions can create more consistent, accurate and useful calibration of employee contributions. It provides raters with a sense of the level of differentiation they are expected to make when assessing employees. For example, if raters know they are expected to place at least 5% and no more than 10% of employees in the highest category they will challenge each other to achieve this target. Raters should be told that it is okay not to meet the expectation, but that if they do not meet it they must provide a clear explanation why their group of employees is so different from the other groups being assessed in the company.

Two things are required to set expected calibration outcomes. First, you must decide the number of assessment categories to use. Table 4 list some example categories used for performance, potential, and compensation calibration. Note that these titles are just examples. Careful thought should be put into choosing category labels that fit the culture and goals of the organization. In addition to the titles, it is also important to define the criteria that should be used to determine which category to place employees. Research on performance evaluations suggests that five categories tend work best for performance calibration[35]. Many companies use three categories when calibrating for potential. The categories used to calibrate compensation decisions tend to vary based on whether the discussion is focused on merit increases, variable pay, or long-term incentives.

In addition to the number of assessment categories, you must decide on the percentage of employees expected to fall in each category. Historically, companies tended to use percentages that follow a normal or near normal distribution. For example, for five performance categories the expectation might be 10% high performer, 25% above expectations, 30% meets expectations, 25% below expectations, 10% serious concerns. Research has shown that employee performance rarely follows these sorts of distributions37, with far fewer people falling in the upper and lower categories than would be expected based on a strict normal curve. Based on this, the following might be a more realistic performance distribution: 5% high performer, 40% above expectations, 35% meets expectations, 15% below expectations, and 5% serious concerns. In addition, if a company actively addresses

Table 4. Example categories used for performance and potential calibration

Performance Categories	Potential Categories	Compensation Categories
Outstanding	Outgrown current role;	Maximum increase possible
Exceptional	time for job transition	Above average increase
Meets Expectations	Actively developing for a future transition	Average increase
Below Expectations		Below average increase
Serious Concern	No imminent role change expected	No increase
Critical Talent		
High Value	Ready now for promotion	
Solid Contributor	Ready in 1 to 3 years	
Struggling	Well placed	
Seriously Underperforming		

performance issues, it may have no employees in the lowest category since they will have been actively managed up or out of the organization.

The number and type of categories used for calibration and the expected distributions across these categories should be tailored to match the nature of the work, business culture, and talent management philosophies of your company. What is important is not so much how many rating categories you have or what percentages are expected to fall into each category. What is important is that the categories and percentages make sense to the managers and employees involved in the calibration process.

- **The use of 9 box assessments for calibration.** The "9 box" is a very common categorization method used for calibration. For this reason alone, we feel it is important to discuss its advantages and disadvantages. The concept of the 9 box is fairly simple. Employees are placed in a three by three grid containing nine different boxes based on how they are assessed on two dimensions. Figure 6 shows the most common version of the 9 box. This version is used for succession management with employee placement based on ratings of current job performance and future job potential. Another common version of the 9 box used for performance management places employees based on goal accomplishment (what employees have achieved) and performance behavior (how they achieved it). Nine boxes are usually configured so the strongest employees are in the upper right box and the weakest are in the lower left.

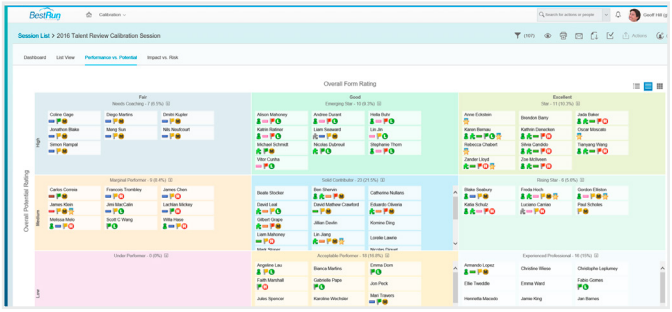


Figure 6. Typical 9 Box Used for Succession Management

The advantage of 9 boxes is they reinforce that employee value is a multidimensional concept. Success depends on different, independent characteristics. For example, a high performing engineer in a technical individual contributor role may not necessarily have the potential to be a great engineering manager. And a sales person who exceeds their sales targets is not a truly high performing employee if they mistreated their colleagues or misled customers to achieve their goals.

The disadvantage of 9 boxes is they often create confusion and resentment among managers and employees. People struggle to differentiate between the two dimensions used to place people on a 9 box because "performance & potential" or "what & how" tend to overlap in the actual world. Employees who exhibit extremely high performance and very low potential are rare in most jobs (sales and highly technical roles possibly being an exception). Similarly, there are few jobs where people can achieve all the right things if they are doing things all the wrong way. Because performance tends to correlate with potential and what people

achieve usually depends on how they achieve it, it is very rare to find employee who strongly fit the upper left or lower right corners of a 9 box. These types of employees do exist, but they are fortunately relatively rare (Bob Sutton's book "The no asshole rule" provides an excellent discussion of the dangers posed by people who achieve the right things the wrong way). Because managers and employees rarely see people who clearly belong in the upper left or lower right corners of a 9 box, these 9 box categories don't reflect the reality of the world as it is experienced by most people. This makes the 9 box process feel artificial and frustrating for many managers and employees.

Managers also struggle to communicate the meaning of 9 box assessments to employees. Many of the 9 box categories don't make sense to managers or employees. It is clear that being in the upper right of a 9 box is better than being in the middle, and being in the lower left is bad. But what does it mean to be in the boxes off the diagonal? Is the middle top box better than the middle right box? Managers already struggle to differentiate between "performance vs. potential" and "what vs. how" during the 9 box assessment process. They struggle even more when asked to explain 9 box results to employees who are concerned about what their placement in different boxes means for their careers.

We are seeing companies replace 9 boxes with calibration talent reviews that do not try to look at performance and potential independently. Instead they focus on categorizing employees along a single dimension such as:

- **Impact:** Which employees are having a disproportionate impact on the organization's success, and how should the company manage people differently based on the value they are providing?
- **Investment:** How should the company invest limited financial resources such as compensation in a manner that will maximize future workforce productivity?
- **Advancement:** How can the company most effectively support the career growth, retention and development of employees possessing the leadership potential and/or critical expertise needed to support future business goals?

These calibration sessions focus on very specific topics and are clear about what they are not addressing. For example, the discussion of impact only focuses on performance in the current role. It does not address what compensation should be given to people nor does it rate people on potential for future roles. This allows them to use simplified assessment models where employees are placed into 3 to 5 categories aligned on a single dimension (as opposed to the 9 category, two-dimensional model of a 9 box).

7. WHO WILL PARTICIPATE IN CALIBRATION SESSIONS AND WHAT ARE THEIR ROLES?

A particularly critical decision when designing calibration methods is who should be included in the group of raters responsible for assessing the employees. It is surprising how often companies base this decision solely from organizational charts without thinking through important factors such as people's level of exposure to the employees being rated, their perspective

on the employees' work and its impact on the business, their understanding of the job and company culture, or their experience and training related to assessing performance or potential. These questions are all worthy of attention.

Considerable research has been conducted examining how the composition of groups affects the decisions they make. The following research findings are particularly relevant considerations for forming calibration rating groups:

- **Group size.** The size of a decision-making group can influence the effectiveness of its decisions. Compared to small groups (e.g. less than 5 people), large groups will experience more conformity, decreased individual participation from group members, and increased coordination problems and areas of potential conflict[40, 41, 42]. Some research suggests that once a group has more than 7 members, each additional member reduces ultimate decision effectiveness by 10%[43]. While there are no hard and fast rules regarding group size, it is probably best to have somewhere between 3 and 10 raters involved in a calibration session.
- **Shared biases or perspectives.** One of the advantages of calibration is its ability to reduce the risk of biased assessments. However, group decision making can also amplify biases if members of the group share the same bias[23,24]. Shared biases can result from group members having similar demographic characteristics, experiences, or positions. As we have a bias for people who are "like us", we tend to rate similar individuals more positively than less similar individuals[44]. For example, both black and white raters have been shown to give significantly higher performance ratings to members of their own race[45, 46]. However this bias can also extend to reflect a preference for people of the same gender, age, and even organizational role or tenure[47].

  Diversity amongst the raters in a calibration group can play a critical role in achieving fair and unbiased evaluations. Research suggests that increasing within-group diversity can also offer other benefits, such as more productive discussions and higher quality ideas[16, 48]. Consequently, one of the things to consider when creating calibration rating groups is whether to include certain people intentionally because they are not like other members of the group. This is also an argument for including raters and/or facilitators from other parts of the organization who have not had any direct exposure to the employees or jobs being assessed. By participating in the conversation, these raters can help ensure that the assessment decisions being made by the group are in line with how assessment decisions are being made in other parts of the company.

- **Previous experience rating and/or promoting employees.** It is likely that some calibration raters will have previously rated employees included in the calibration session before. For example, a manager may rate the same employees who report to him/her several years in a row. While this may be unavoidable, it poses risks worthy of consideration. Research shows that when a rater has evaluated an employee in the past, they are more likely to discount new information about this employee if the information conflicts with their original evaluation[49]. In other words, raters may exhibit a confirmation bias to interpret new information in a way that confirms our preexisting beliefs while giving disproportionately less consideration to

alternative possibilities[50]. It may therefore be wise to include raters in the session who have not previously evaluated any of the employees.

Studies have also shown that raters tend to evaluate employees more favorably if they had previously hired or promoted that employee[49]. This is an example of the sunk cost effect which is a tendency of individuals to make decisions for the sake of justifying a previous decision[49]. Research suggests that groups actually increase the severity and frequency of sunk costs[51], meaning that a manager who hired or promoted an employee may be reluctant to rate that employee poorly in front of their peers, as it might be perceived as an admission to having made a bad staffing decision.

- **Group facilitator.** A group facilitator is "a person whose selection is acceptable to all members of the group, who is substantively neutral, and who has no substantive decision-making authority diagnoses and intervenes to help a group improve how it identifies and solves problems and make decisions, to increase the group's effectiveness" (Schwarz, 1994, p.4). Research suggests that having a facilitator can improve the functioning and performance of small groups[52]. Having a neutral party present during what can become an intense conversation may be invaluable to the success of calibration sessions. Responsibilities our customers commonly assigned to group facilitators include:
  - Pre-work including scheduling the sessions, gathering any supplementary materials needed during the sessions, preparing session agendas, etc.
  - Preparing leaders for any difficult discussions/decisions they foresee.
  - Facilitating the actual conversations including challenging raters to explain their decisions to ensure they are critically and fairly evaluating employees based on appropriate criteria.
  - Providing general support and guidance to help the group work through difficult issues.
  - Resolving conflict and deescalating emotions that may flare up when having difficult discussions about the relative contributions made by employees that raters may know well at both a personal and professional level.

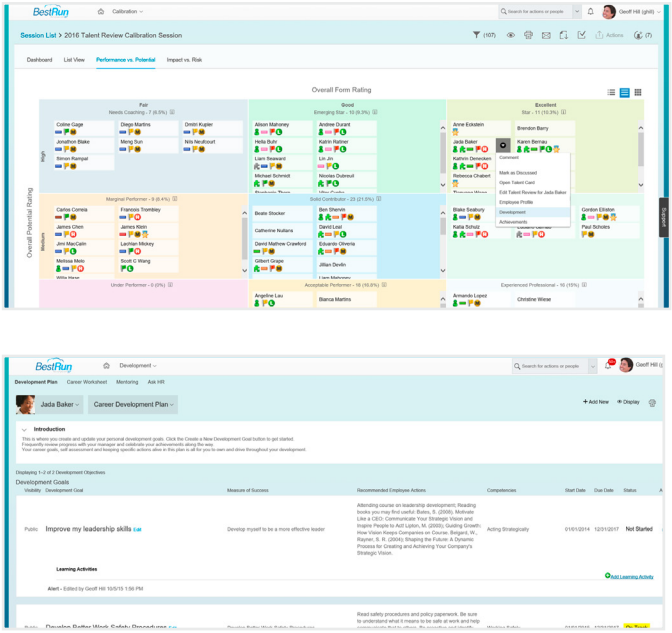


Figure 7: Technology illustration of calibration session facilitation tools. This screen shot shows a tool that allows companies to drill down through different levels of employee information during a calibration session. In the example, an employee is identified in a session, their name is clicked on to show a range of attributes, one of the attributes titled "development plan" is clicked on to show employee progress and recent achievements.

"We find having a neutral facilitator to be hugely beneficial. Leaders in these groups can become heated and passionate sometimes. But facilitators don't have these pre-perceptions or judgments. Facilitators can act as an objective set of ears that help to guide the conversation."

The decisions made by a group are only as good as the members of that group. Keep this in mind when deciding who to include in calibration sessions. Consider factors such as similarity to other raters, previous experience with employees included in the session, and group size. If possible, have a trained facilitator involved in each calibration session. Avoid basing the decision of which raters to include using the organizational chart alone. Ideally take steps to ensure calibrations session includes raters with different perspectives and backgrounds. And put processes in place to ensure the views of all raters are effectively considered and addressed during the discussion.

- **Ensure calibration participants understand and can perform their roles.** The people who participate in calibration sessions tend to fall into three different roles:
  - **Rater:** these are the people who must decide how to categorize the employees being assessed during the session. Raters are usually the managers of the employees being calibrated. They may also include other individuals brought into the calibration session whose perspectives are felt to be valuable and relevant to accurately assessing the employees.
  - **Facilitator:** this is a neutral person whose job is to ensure the calibration session runs smoothly and effectively. These are often HR professionals, but can also be trained employees from other parts of the organization or 3rd party consultants. Facilitators guide the rating process, but typically do not make ratings themselves.
  - **Senior leader:** this is a person higher up in the organization who is attending the session to support and listen to the discussion. The senior leader is often the person who the raters themselves report up to.

Part of designing a calibration process should include defining and guiding people on how to effectively perform these three roles. The following are a few things to keep in mind when doing this:

- **Rater effectiveness:** Research has shown that raters who receive training on how to assess employees are far more effective than those who have not been trained[53]. Rater training should ideally include:
  - An overview of the company's calibration philosophy and expected goals and outcomes.
  - A description of the calibration process and roles of session participants.
  - A description of common rater biases and how to avoid them.
  - "Frame of reference" training where raters complete a few rounds of practice ratings and then receive feedback on their accuracy based on comparisons to standardized "true" ratings developed by the organization[54].

Most of the customers we interviewed for this research study offer some type of rater training, but it was not always mandatory. Some customers argued against the use of mandatory rater

training as it could create frustration among managers. The point was made that if managers aren't taking their responsibilities as raters seriously it reflects a larger issues having to do with leadership and organizational culture. And trying to "force" people to learn rarely works.

"If we have managers who aren't willing to do a proper evaluation, mandatory training isn't going to help them. It will only raise resistance. Fixing that problem requires a different process, like selecting new leadership".

- **Facilitator effectiveness.** Group facilitators perform three particularly critical tasks: directing meaningful conversation, emphasizing rater accountability, and providing conflict resolution.
  - **Directing meaningful conversation** requires facilitators to take an active role in the group's discussion. Facilitator should encourage group members to think critically, provide specific examples and justify their ratings in a productive and respectful way. The following facilitation techniques have been shown to be valuable during group decision making meetings such as calibration sessions[55, 56, 57.] 'Devil's advocacy' involves assigns the role to one group member (or facilitator) to find as many faults and objections to proposed solutions and alternatives as possible[58, 59]. Devil's advocacy can mitigate 'groupthink'[60], and foster analytical skills, communication skills, and emotional intelligence[61]. **'**Dialectical inquiry **'** is used to encourage debate between two opposing viewpoints[55] by noting the benefits and limitations of each set of ideas, stimulating conflict in a constructive way[63]. This strategy can also control the effects of groupthink[60]. Rating justification requires leaders to provide specific examples and explanations to support their chosen ratings. Asking raters to justify their ratings may result in more careful consideration of performance behaviors[64].
  - **Emphasizing rater accountability** requires facilitators to highlight rater responsibilities during the calibration process. Research has shown that emphasizing accountability increases rater motivation[65]. For example, forewarning rater groups that the accuracy of their assessments would be evaluated by others led to greater deliberation and information sharing within the group[34].
  - **Conflict resolution** can be critical in calibration sessions where raters may have significantly different opinions regarding employees. Conflict may result from differences in viewpoints or preferences[66], personality attributes[67], group size and cohesiveness[68], or time limitations[69]. To be effective at conflict resolution, facilitators must be able to diagnose the problem, generate a solution, and enact this solution within the group quickly and efficiently[68]. They must also be perceived to be an objective, neutral party. This is a major reason why facilitators should not also be raters.

- **Senior leader effectiveness.** Senior leaders have one of most delicate and challenging roles during a calibration session. On one hand, they are responsible for ensuring the results of the session meet company expectations regarding accurate assessment of employees and expected distribution of ratings. On the other hand, they must be careful not to come across as "telling managers how to rate" or taking other actions that lead managers to feel a loss of ownership over the assessment decisions. The senior leaders must be responsible for the total

calibration session results, but the raters must be responsible for the individual employee assessments. Senior leaders should be made aware of the delicate nature of their role. The facilitator and senior leader will ideally collaborate closely to challenge and coach raters on how to make assessment decisions while being careful not come across as telling raters what assessment decisions to make.

Few decisions have a greater long-term impact on company culture and success than compensation and staffing decisions tied to employee performance and potential. It is surprising how many companies allow managers to make these decisions on their own with little to no clear guidance and training. Calibration methods can greatly improve workforce management by ensuring managers are accurately assessing employee contributions and potential. But calibration will not work well unless raters are guided on how to effectively discuss and accurately assess employees. Simply raising awareness of common assessment biases and encouraging raters to avoid committing them is not enough[54,70]. Accurate assessment requires providing raters with feedback on the accurate of their ratings. It requires facilitators to play an active role in challenging calibration session members including those who may be more senior to them on the organizational chart. Last, it requires active involvement from senior leaders who can guide raters to make accurate assessments of employees without crossing the line into telling them how to rate employees.

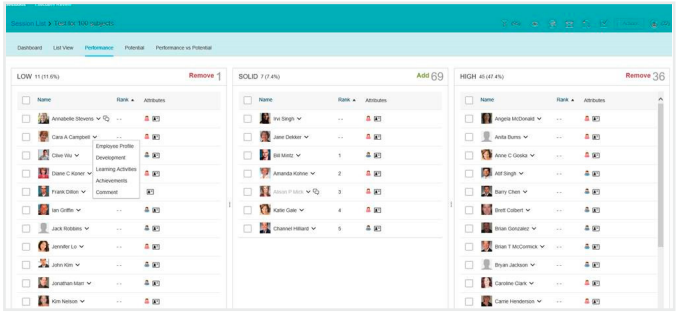


Figure 8: Technology illustration for training embedded into a calibration process. In this example, a short training program on performance assessment is integrated into the tools used by managers to conduct calibration sessions. Incorporating "just in time" learning resources into talent management process technology is often viewed as more effective than training programs that are conducted outside of the process.

8. HOW WILL THE RESULTS OF CALIBRATION SESSIONS BE COMMUNICATED?

Calibration methods can increase employee engagement and satisfaction provided they are effectively designed, consistently followed, and appropriately communicated. Employee assessment processes that are deemed accurate and unbiased by employees have been shown to create greater satisfaction with appraisals overall, increased job satisfaction and commitment, increased productivity, and decreased intention to quit[71]. But what makes employees perceive an assessment process to be accurate and unbiased? It starts with having a well-defined and consistently applied assessment process. But the process itself is not

enough; it is just as important to explain the process and its results to employees in a manner that is felt to be fair, accurate and appropriate. This requires transparency around calibration process, criteria, and philosophy.

"We believe it is extremely important to have transparency. People become frustrated and suspicious if they have a feeling that something is done in a black box behind their backs, and it can lead to de-motivation".

The way information is presented has a major impact on how people interpret and react to it. There are specific things employees can be told before, during, and after calibration that will improve their feelings of fairness and overall sentiment associated with calibration. These things are directly related to organizational justice, or employees'perceptions of fairness in the workplace[72]. Organizational justice can be broken down into three main categories, interpersonal justice, procedural justice, and distributive justice. The following is a discussion of each of these types of justices, along with examples of methods customers have used to support them.

- **Interpersonal justice:** This is associated with how information and decisions are communicated in terms of the interpersonal nature of the communication[8]. Interpersonal justice reflects the degree to which employees feel that information has been presented in a caring, sensitive and personal manner; for example, communicating assessment results through in-person conversations as opposed to sending results to employees through blanket e-mails. Interpersonal justice influences employee satisfaction with performance rating results, personal feelings toward company managers and leaders, and commitment and effort toward work[71].

  The key to interpersonal justice is to focus on how information is communicated. Are managers trained on how to deliver assessment results in a sensitive and positive manner? Are they held accountable for meeting with employees in person in a timely manner? Do they take time for employees to express their feelings and reactions?

  "Included in our training programs is a workshop on feedback and coaching which trains leaders on how to give feedback to employees, as well as a 'review' session on how to have that final conversation with employees and to summarize results in a meaningful way to look forward to the next year."

- **Procedural justice:** This is associated with the procedures and policies used to determine ratings[74]. Procedural justice is achieved when employees understand the methods used to evaluate their performance and trust that these methods are fairly and consistently applied. Procedural justice influences work performance[75], pay satisfaction[75,] appraisal system satisfaction[7], trust in management[7], and organizational commitment[76]. The following are a few things our customers have done to support procedural justice:
  – Make the rating criteria used by leaders during calibration process public and available to all employees. Remember, employees know that they are going to be evaluated in some manner. The more they know about how they are evaluated the more they can control their own careers.

– Give employees an opportunity to self-assess themselves and then discuss the accuracy of their perceptions. Research has shown that use of self-assessment data increases employee motivation and decreases levels of concern about unfair treatment[77, 78]. At the same time, employees often overestimate their skills and are overconfident in their judgments[79]. So part of the self-assessment process should include a learning step to increase employee self-awareness. This step will be more effective if you ask employees to provide detailed examples and justifications for a good rating, rather than simply giving themselves an overall rating.
– Post a calendar and flow chart so employees can see when different steps in the calibration process are conducted and how data from calibration is used to make organizational decisions. Provide a forum for employees to ask questions to better understand the company's calibration process and philosophy. The goal of this is not to get all employees to enthusiastically endorse the process (this will probably never happen!). It is to make sure that employees understand the company's culture, values and beliefs and the rationale behind them.

"We really do believe in transparency here. We want our employees to know we're not trying to keep secrets. If you find you were evaluated a certain way, we want you to know and understand how that came about. The more we explain what happens, the less speculation and suspicion there will be on the other end."

- **Distributive justice:** This focuses on fairness associated with the assessment result and whether people feel they have received the "right" rating[80]. Distributive justice does influence employee satisfaction with the assessment process[7], but it has far less impact on overall employee attitude and performance than procedural and interpersonal justice. Most people can accept that they may not always receive the highest ratings, provided they believe the process was appropriately designed, clearly communicated, consistently and fairly followed, and appropriately and sensitively communicated. However, companies should be prepared for some level of negative reaction when people receive lower ratings.

There are several things companies can do to ensure greater perceptions of distributive justice. First, provide employees an opportunity to express their opinions and ask questions about the assessment results. Sincerely listening to employee concerns can increase perceptions of justice even if the assessment results remain unchanged. Second, avoid surprises by making sure managers are providing ongoing performance feedback to employees throughout the year. Third, focus on future opportunities not past issues. Emphasize ways to improve and stress that ratings can and do change over time. Express appreciation for employee contributions and confidence in their future success.

"We're talking about human beings. Their responses depend very much on the ratings they receive. If it's a good rating, people become very interested to learn more about the process and what people have said about them. When it's a bad rating, the person will always question the rating and how it came about".

The purpose of calibration is not to "document the past" but to influence the future. The primary way calibration influences the future is through increasing the accuracy of organizational workforce decisions and positively impacting the attitudes, motivation, development, and careers of individual employees. Much of the impact of calibration hinges on how calibration processes results are communicated to employees. Calibration processes do not end with the calibration sessions. It is critical to think through how calibration results will be communicated to employees afterwards, as this will have the greatest impact on organizational performance. Calibration design includes having a clear process to communicate assessment results to employees in a manner that emphasizes organizational justice and is felt to be fair, accurate and appropriate.

"The only reason we assess past performance is because it is the best way to predict and influence future performance."

Conclusion

Over the past several years, there has been a trend in human resources calling to "abolish ratings". Performance ratings, some people argue, are inaccurate and create negative reactions in employees. While this movement is well intentioned, it is also misguided. The act of rating is about placing employees in different categories based on their perceived value to the organization. If your company believes that not everyone performs at the same level, and that employees who contribute more to the company should be rewarded with greater resources and given greater career opportunities within the organization, then your company rates employee performance in some manner. Your company may not use annual performance ratings or assign numeric ratings to employees, but leaders in your company are using some method to categorize certain employees as being more valuable than others. The question is not whether you rate employees, but whether you rate them in an accurate, fair and effective manner.

We believe that calibration provides one of the most promising methods for improving the accuracy and value of performance management, succession management, compensation, and talent management overall. Calibration processes, when done well, help ensure that high performing employees who contribute relatively more to the company receive greater levels of investment in the form of pay and career growth opportunities, while low performing employees are identified, supported and if necessary, held accountable for improving their performance. The conversations and data created through calibration provide insight and shared understanding between employees, managers and leaders on what employee behaviors, skills and attributes are important to the organization, why they are important, and how to effectively develop and retain the employees who possess them. This sort of shared understanding is what it means to have a true high performance culture.

The concept of calibration is not new. People have been meeting to discuss team member contributions probably as long as people have been working in teams. Calibration conversations are a very natural human activity. But like all human conversations, the value of calibration session depends on how they are conducted. This paper summarizes best practices gathered from SAP SuccessFactors customers with insights from organizational psychology research to help companies maximize the value of calibration as tool to create more engaged, productive, and fair work cultures. We hope that this paper has provided you with some useful tips and insights to gain the most value from calibration for your organization.

REFERENCES

1. Viswesvaran, C. & Ones, D. (2000). Perspectives on models of job performance. International *Journal of Selection and Assessment*, *8*, 216-228.
2. Boudreau, J. Ramstad, PM (2005). Talentship, talent segmentation, and sustainability: A new HR decision science paradigm for a new strategy definition. *Human Resource Management*, *44*, 129-136.
3. O'boyle, E., & Aguinis, H. (2012). The best of the rest: revisiting the norm of normality and individual performance. *Personnel Psychology*, *65*, 79-119.
4. Bloom, N., & Van Reenen, J. (2007). Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics*, *122*, 1341–1408.
5. Colquitt, J., Conlon, D., Wesson, M., Porter, C., & Yee, K. Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, *Vol 86*(3), 425-445.
6. Auginis, H., Joo, H. & Gottfredson, R.K. (2011). Why we hate performance management – and why we should love it. *Business Horizons*, *54*, 503-507.
7. Austin, J,T. & Villanova P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology*, *77*, 836-874.
8. Elliot, Andrew J. (Ed), Dweck, Carol S. (Ed). (2005). *Handbook of competence and motivation*, (pp. 566-578). New York, NY.
9. Hastings, R. (2012). Most large companies calibrate performance. *HR Magazine*, *87*.
10. Aguinis, H. (2009). Performance management. Prentice Hall.
11. Holtom, B., Mitchell, T., Lee, T., & Eberly, M. (2008). Turnover and Retention Research. *The Academy of Management Annals*, *2*(1), 231-274.
12. Steers, R., & Mowday, R. (1981). Employee turnover and post-decision justification. In L.L. Cummings & B.M. Staw (eds.), *Research in organizational behavior*. Greewich, CT: JAI Press.
13. Wright, T., & Bonnett, D. (1992). Job satisfaction and psychological well-being as non-additive predictors of workplace turnover. *Journal of Applied Psychology*, *3*, 486-493.
14. Wright, T., & Bonnett, D. (2002). The moderating effects of employee tenure on the relation between organizational commitment and job performance: a meta-analysis. *Journal of Applied Psychology*, *87*(6), 1183-1190.
15. Scullen, SE., Mount, MK, & Goff, M (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, *85*, 956-971.
16. Kolbe, M., & Booz, M. (2009). Facilitating group decision-making: Facilitator's subjective theories on group coordination. *Forum: Qualitative Social Research*, *10*(1).
17. Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, *29*(1), 80-85.
18. Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, *29*(1), 80-85.

19. Schleicher, D., Bull, R., & Green, S. (2009). Rater reactions to forced distribution rating systems. *Journal of Management*, *35*, 900-930.
20. Scullen, S., Bergey, P., & Aiman-Smith, L. (2005). Forced distribution rating systems and the improvement of workforce potential: A baseline simulation. *Personnel Psychology*, *59*, 1-32.
21. Bates, S. (2003). Forced ranking. *HR Magazine*, p. 62-68.
22. Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *The American Journal of Psychology*, *60*(1), 1-29.
23. Argote, L., Devadas, R., & Melone, N. (1990). The base-rate fallacy: Contrasting processes and outcomes of group and individual judgment. *Organizational Behavior and Human Decision Processes*, *46*, 296-310.
24. Martell, R., & Borg, M. (1993). A comparison of the behavioral rating accuracy of groups and individuals. *Journal of Applied Psychology*, *78*(1), 43-50.
25. Landy, F., & Farr, J. (1983). The measurement of work performance: Methods, theory, and applications. New York: Academic Press.
26. Scott, D., McMullen, T, & Royal, M. (2011). Reward fairness: Slippery slope or manageable terrain? *WorldatWork*. https://www.worldatwork.org/waw/adimLink?id=53154
27. Bono, E. (1999). Six Thinking Hats. Back Bay Books.
28. Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making. *Journal of Personality and Social Psychology*, *48*, 1467-1478.
29. Janjua, S. (2012). 7 best practices for effective group decision making. *Philosophy IB.* (http://www.philosophyib.com/3/whole-brain/effective-group-decision-making)
30. Baumeister, R. F. (2003). The psychology of irrationality: Why people make foolish, self-defeating choices. In I. Brocas & J. Carillo (Eds.) T*he Psychology of Economic Decisions*. New York: Oxford University Press.
31. Tierney, J. (2011). Do you suffer from decision fatigue? *New York Times*. http://www.nytimes.com/2011/08/21/magazine/do-you-suffer-from-decision-fatigue.html
32. Kravitz, D., & Balzer, W. (1992). Context effects in performance appraisal: A methodological critique and empirical study. *Journal of Applied Psychology*, *77,* 24-31.
33. Riege, A. (2005). Three-dozen knowledge-sharing barriers managers must consider. *Journal of Knowledge Management*, *3*, 18-35.
34. Martell, F., & Leavitt, K. (2002). Reducing the performance-cue bias in work behavior ratings. Can groups help? *Journal of Applied Psychology*, *87*(6), 1032-1041.
35. Eichinger, R.W., Lombardo, M.M., & Ulrich, D. (2006). 100 things you need to know: best people practices for managers & HR. Minneapolis, MN: Lominger.
36. Fasold, F., Memmert, D., & Unkelbach, C. (2015). A theory-based intervention to prevent calibration effects in serial sport performance evaluations. *Psychology of Sport and Exercise*, *18*, 47-52.
37. O'Boyle Jr., & Aguinis, H. (2012). "The Best & the Rest: Revisiting the Norm of Normality of Individual Performance" *Personnel Psychology*, *65*, 79-119

38. Bell, S. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, *92*(3), 595-615.
39. Cawley, B., Keeping, L., & Levy, P. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, *83*(4), 615-633.
40. Asch, S. (1951). Effects of group pressure on the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men*. Pittsburgh, PA: Carnegie Press.
41. Hare, A.P. (1952). Interaction and consensus in different sized groups. *American Sociological Review*, *17*(2), 261-267.
42. Steiner, I.D. (1972). Group processes and productivity. New York: Academic Press.
43. Blenko, M., Rogers, P., & Mankins, M. (2010). Decide and deliver: Five steps to breakthrough performance in your organization. Bain & Company, Inc.
44. Byrne, D. (1971). The attraction paradigm. San Diego: Academic Press.
45. Kraiger, K., & Ford, J.K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology*, *70*, 56-65.
46. Landy, F., & Farr, J. (1980). Performance rating. *Psychological Bulletin*, *87*, 72-107.
47. Milliken, F., & Martins, L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review*, *21*(2), 402-433.
48. Schwenk, C., & Cosier, R. (1993). Effects of consensus on devil's advocacy on strategic decision-making. *Journal of Applied Psychology*, *23*(2), 126-139.
49. Bazerman, M., Beekun, R., & Schoorman, F. (1982). Performance evaluation in a dynamic context: A laboratory study of the impact of a prior commitment to the ratee. *Journal of Applied Psychology*, *67*(6), 873-876.
50. Plout, S. (1993). *The Psychology of Judgment and Decision Making*. Reno: Sierra Nevada Books.
51. Whyte, G. (1993). Escalating commitment in individual and group decision making: A prospect theory approach. *Organizational Behavior and Human Decision Processes*, *54*, 430-455.
52. Schwarz, R. (1994). *The Skilled Facilitator: A Comprehensive Resource for Consultants, Facilitators, Managers, Trainers, and Coaches* (Revised Edition) Jossey-Bass, 2002, 1-432.
53. Davis et al. (1992) Effects of expert and non-expert facilitators on small-group process and on student performance
54. Podsakoff, 2003
55. Schweiger, D., Sandberg, W., & Rechner, P. (1989). Experiential effects of dialectical inquiry, devil's advocacy, and consensus approaches to strategic decision making. *Academy of Management Journa*l, *22*(4), 745-772.
56. Watson, W., Kumar, K., & Michaelsen, L. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal*, *36*(3), 590-602.

57. Li, J., & Karakowsky, L. (2001). Do we see eye-to-eye? Implications of cultural differences for cross-cultural management research and practice. *The Journal of Psychology*, *135*(5), 501-517.
58. Schwenk, C. (1984). Devil's advocacy in managerial decision-making. *Journal of Management Studies*, *21*(2), 153-168.
59. Schweiger, D., & Finger, P. (1984). The comparative effectiveness of dialectical inquiry and Devil's advocacy: The impact of task biases on previous research findings. *Strategic Management Journal, 5*(4), 335-350.
60. Lunenburg, F. (2012). Devil's advocacy and dialectical inquiry: Antidotes to groupthink. *International Journal of Scholarly Academic Intellectual Diversity*, *14*(1).
61. Kreitner, R., & Kinicki, A. (2010). Organizational behavior (9th ed.). New York, NY: McGraw-Hill.
62. Katzenstein, G. (1996). The debate on structured debate: Towards a unified theory. *Organizational Behavior and Human Decision Processes*, *66*(3), 316-332.
63. Giumetti, G., Schroeder, A., & Switzer, F. (2015). Forced distribution rating systems: When does "rank and yank" lead to adverse impact? *Journal of Applied Psychology*, *100*(1), 180-193.
64. Williams, K.J., DeNisi, A.S., & Cafferty, T.P. (1985). The role of appraisal purpose: Effects of purpose on information acquisition and utilization. *Organizational Behavior and Human Performance*, *35*, 314-339.
65. Bretz, R.D., Milkovich, G.T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions and implications. *Centre for Advanced Human Resource Studies*.
66. McGrath, J. (1984). Groups: Interaction and performance. Edgewood Cliffs, NJ: Prentice-Hall.
67. Baron, R. (1989). Personality and organizational conflict: Effects of the type A behavior pattern and self-monitoring. *Organizational Behavior and Human Decision Processes*, *44*(2), 281-296.
68. Farmer, S., & Roth, J. (1998). Conflict-handling in work groups: Effects of group structure, decision processes, and time. *Small Group Research*, *29*(6), 669-713.
69. Baxter, L. (1982). Conflict management: An episodic approach. *Small Group Behavior*, *13*(1), 23-42.
70. Woehr, D., & Huffcutt, A. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *64*, 189-205.
71. Taneja, S., Srivastava, R., & Ravicandran, N. (2015). Consequences of performance appraisal justice perception: A study of Indian banks. *The IUP Journal of Organizational Behavior*, *14*(3), 33-59.
72. Greenberg, J. (1987). A taxonomy of organizational justice theories. *Academy of Management Review*, *12*(1), 9-22.
73. Smither, J.W. (1988). Lessons learned: Research implications for performance appraisal and management.
74. Greenberg, J. (1986). Determinants of perceived fairness of performance evaluations. *Journal of Applied Psychology*, *64*, 349-371.
75. Swiercz, P.M., Icenogle, M.L., Bryan, N.B., & Renn, R.W. (1993). Do perceptions of performance appraisal fairness predict employee attitudes and performance. *Academy of Management*.

76. Mcfarlin, D.B., & Sweeney, P.D. (1992). Distributive and procedural justice as predictors of satisfaction with personal and organizational outcomes. *The Academy of Management Journal*, *35*(3), 626-637.
77. Roberts, G.E. (2003). Employee performance appraisal system participation: A technique that works. *Public Personnel Management*, *32*(1), 89-98.
78. Kromrei, H. (2015). Enhancing the annual performance appraisal process: Reducing biases and engaging employees through self-assessment. *Performance Improvement Quarterly*, *28*(2), 53-64.
79. Dunning, D., Heath, C., & Suls, J.M. (2004). Flawed self-assessment implications for health, education, and the workplace. *Psychological science in the public interest*, *5*(3), 69-106.
80. Folger, R. (1977). Distributive and procedural justice: Combined impact of voice and improvement on experienced inequity. *Journal of Personality and Social Psychology*, *45*, 108-119.

ABOUT SAP SUCCESSFACTORS

SAP SuccessFactors provides the leading cloud-based HCM suite, which helps HR drive business execution with solutions that are complete, beautiful and flexible enough to start anywhere — optimizing your workforce today and preparing them for tomorrow. SAP SuccessFactors solutions are supported by a global partner ecosystem and the experience and commitment of SAP. With approximately 45 million subscription seats globally, we strive to delight our customers by delivering innovative solutions, content and analytics, process expertise and best practices insights from serving our broad and diverse customer base. Today, with more than 6,000 customers, we serve organizations of all sizes across more than 60 industries in more than 177 countries using our application suite in 42 languages.

ABOUT SAP

As market leader in enterprise application software, SAP (NYSE: SAP) helps companies of all sizes and industries run better. From back office to boardroom, warehouse to storefront, desktop to mobile device — SAP empowers people and organizations to work together more efficiently and use business insight more effectively to stay ahead of the competition. SAP applications and services enable more than 253,500 customers to operate profitably, adapt continuously, and grow sustainably. For more information, visit www.sap.com.

SAP SuccessFactors

www.sap.com/contactsap

SAP

Run Simple