

Unring the Bell Curve of Performance Evaluations

“Drive Differentiation through Innovation”

Rico Martino
June 9, 2015

Overview

The practice of using the bell curve, or normal distribution, in performance evaluations is both widespread and controversial. This paper presents a reasonable alternative by drawing upon a widely accepted process of performance measurement in one field to apply to the field of corporate management. In some ways, the approach outlined here is simpler, in other ways more involved. Overall, this paper contends, this approach is fairer and will lead to greater acceptance by the employees being evaluated.

Summary

This paper proposes that performance evaluations contain these eight key points:

1. Preconceived assumptions about the distribution of human performance are irrelevant.
2. Use absolute ratings not relative rankings.
3. Use decimals not integers.
4. Evaluate all employees on five common characteristics and one overall rating.
5. Each of the five common characteristics has a predefined rubric for each job title.
6. The relative weights of the five common characteristics that yield the overall rating are averaged across the corporation.
7. Publish historical precedents.
8. Ask the employees before implementing.

History

For the past four years at least, Raymond James Financial has been using the bell curve as a basis for their annual performance evaluations. The assumption that human performance is distributed according to the bell curve, or normal distribution, is not supported by any peer-reviewed research papers. The continued use of false assumptions in corporate management lessens management’s credibility.

The good news is that [Raymond James engenders a culture of valuing diversity](#)¹ and has an excellent [Tuition Assistance Program](#).² The author earned his MBA in finance and management through their Tuition Assistance Program and is now applying his education to help the company that helped him.

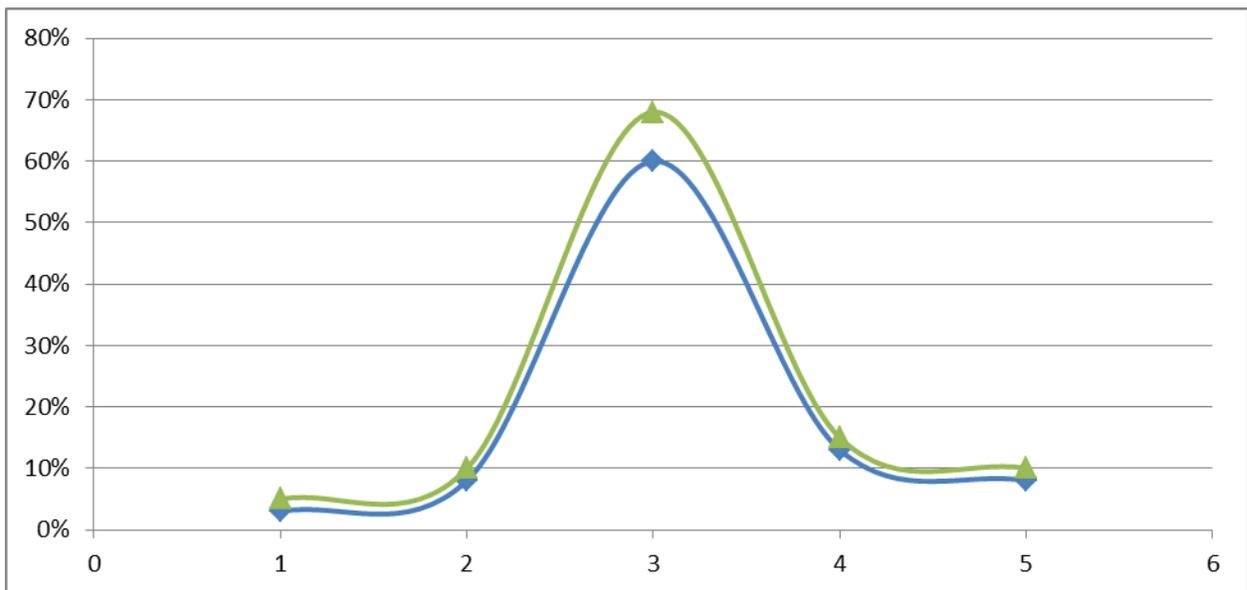
Raymond James prides itself on its innovations in technology.^{3,4} Innovation is so important to Raymond James that they include it in their [Mission Statement](#): “Innovation is requisite to our survival in a changing world.”⁵ One of the goals in our Technology and Operations Vision is to “Drive differentiation through innovation.” This paper presents innovations in corporate management.

The Bell Curve

In September 2014, each performance evaluation template at Raymond James included this chart:

Rating	Description
5 – Top Achiever (Top 8% - 10% of population)	Consistently performs at a superior level resulting in significant and impactful results. Contributions far exceed Valued Contributors.
4 - High Contributor (13% - 15% of population)	Consistently performs at a higher level to achieve outcomes. Contributions exceed Valued Contributors.
3 - Valued Contributor (60% - 68% of population)	Fulfills responsibilities of job to consistently meet established goals and responsibilities. A competent and effective performer.
2 - Needs Improvement (8% - 10% of population)	Falls below standards and does not consistently meet established goals. Further improvement is needed to become a competent and effective performer.
1 – Unsatisfactory (3% - 5% of population)	Far below required level of competence. Immediate and significant improvement is needed to meet minimum standards.

Using Excel to graph the above values, we obtain:



Whether the rating number (one through five) is mandated or encouraged, the results are the same: the resultant data will fit the expectations.

Let us say that Gallup, Rasmussen or Zogby told its pollsters before going out to gather statistics: “We expect these results: Clinton 20%, Christie 15%, Rubio 10%, Paul 5%.” What would their results look like? Quite similar to the expectations of their superiors. Of course, the polling companies would not prejudice their employees, and neither should we.

Minimize Observer Bias

In statistics bias is defined as, “Any tendency which prevents unprejudiced consideration of a question.”⁶ Observer bias is caused by “those involved in designing, administering and/or analyzing an experiment or other data collection exercise, influence the results, typically unintentionally.”⁷ By requiring or recommending a bell curve distribution, we are intentionally influencing the results.

Why Be Normal?

The bell curve is more specifically called the normal distribution. The normal distribution is so named because that distribution is normal, or average. It is also called the Gaussian distribution after Carl Friedrich Gauss who described its properties in the early nineteenth century.⁸

There are several other distributions other than the normal distribution, for example, Poisson distribution, power distribution, Maxwell distribution, gamma distribution, Pareto distribution, exponential distribution, geometric distribution, logarithmic distribution, and dozens more. To say that human performance follows the normal distribution without any supporting evidence is perilous.

Most importantly, why concern yourselves with any distribution at all? This paper asserts that you do not. We do not need to assume any preconceived distribution of human performance in order to do fair performance evaluations. Any preconceived distribution is irrelevant.

Continuum of Numbers

The output of the current method of using a bell curve and this proposed method of not using a bell curve both produce the same desired output: a continuum of numbers from 1.0 to 5.0. How these numbers are distributed is irrelevant. That these numbers are different is relevant.

Relative Ranking v. Absolute Rating

Employee evaluation systems that assume some type of preconceived distribution pit one employee against another at the very outset of the evaluation process. This approach is called relative ranking. To illustrate consider this scenario in the words of the author:

“I grew up in a family of five children. If my parents had believed in the bell curve of child performance, we would have had three Average Martinos, one Stellar Martino and one Schmuck Martino. On the day of the Christmas bonus, my parents would have told Schmuck Martino, ‘You can do it. We believe in you. Next year, you can improve to become Average Martino. Of course, one of the Average Martinos will get kicked down to Schmuck Martino, through no fault of his own. But, that’s how things work around here. We believe in the bell curve.’”

Of course, that scenario did not happen, but it happens every year in thousands of corporations to millions of employees. Sadly, employees accept it without question. “How fortunate for leaders that men do not think.”⁹ –Adolf Hitler.

Now, let’s think.

Is there another approach where low performers can improve to average without any consequence to the existing average performer? Can an average performer excel without lowering the rating of another excellent employee?

Yes. It is called absolute rating. Employees are rated against an absolute scale without regard to how others are rated. If 30% meet the Top Achiever status, then 30% get Top Achiever. No curve, no jockeying for position, no ratings changed to meet some preconceived idea of distribution.

Finding Excellent Leaders

Some years ago, there were two baseball teams from the same city, but with significantly different performance records: New York Yankees and New York Mets. The Yankees were one of the most successful sports clubs in the world; whereas, the Mets often finished in the bottom of the standings. If we follow the bell curve or relative ranking philosophy, then the average player on each team would be rated 3.0. This approach does not match our understanding of “Top Performer” and “Needs Improvement”. There must be a better explanation.

Their leaders made the difference.

By using an absolute rating system, excellent leaders now have tangible evidence to support their excellence.

TRIZ

As a preface to the main points of this paper, let us take a quick look at TRIZ, a Russian acronym for “Theory of Inventive Problem Solving.” Principle #3 of TRIZ states, “Creative innovations use scientific effects outside the field where they were developed.”¹⁰ An example is the carcass disassembly line of a slaughterhouse in Chicago providing the idea for an automobile assembly line in Detroit.¹¹

The ideas proposed in this paper originated in academia and statistics. This paper is proposing their adaptation to industry.

Requirements

For any performance evaluation process to be effective, it must meet these requirements:

1. Understandable
2. Fair

It must be understandable to:

1. The manager
2. The employee

It must be fair to:

1. The manager
2. The employee
3. Other employees

How do you know if it meets these requirements? Ask them.

Please note one characteristic that is not a requirement: easy. The process proposed by this paper requires more thinking, but the extra thinking results in fairer evaluations.

How important are your employees, anyway?

GPA

A performance measurement system that most people consider fair is the college grade point average, or GPA. The GPA system is fair because the specifications of each goal and measurements for that goal are approved by an outside, independent accrediting organization.

Each college class has a syllabus distributed on the first day of class. This syllabus itemizes the tasks necessary to achieve the overall goal: to pass the class. The syllabus also details how the final performance rating, i.e., grade, will be determined. For those tasks whose evaluations may be ambiguous, the teacher provides a rubric on how to convert subjective concepts into numerical measurements.

The student then has a clear understanding of what it takes to succeed in that class.

An outside, independent accrediting organization, or a consulting company that is recognized for excellence in human resource management can perform the normalization of levels between disciplines.

Decimals, not Integers

Most colleges give gradations to letter grades, such as A- and B+. An A- equals to 3.67 and a B+ equals 3.33 quality points per credit. Do not be constrained to an integer rating value. A decimal value can be more accurate, plus we have computers to do the calculations.

Five Common Measurements and One Overall Rating

From discussions with managers across several different disciplines, the author has discovered five (5) characteristics common to all performance measurements:

1. Quantity
2. Quality
3. Expertise
4. Helpfulness
5. Initiative

Evaluate each employee regardless of discipline or department on a scale of 1.0 to 5.0 for each of the above characteristics. Use decimals if necessary.

In addition to the five common measurements, the manager assigns an overall rating to the employee's performance. He may use any approach that he wants: from weighted average to gut feeling. These relative weights will be averaged (technically least squares fit) across all other evaluations in the corporation.

Rubric for Each Measurement for Each Job Title

Before the start of the evaluation period, Human Resources (HR) or each overall department shall publish their rubrics on how they will measure each characteristic for each job title.

For example, they will publish what constitutes a 1, 2, 3, 4 or 5 level for quality of an application developer, which will probably be slightly different for a senior application developer. They will publish another rubric for each of quantity, expertise, helpfulness, and initiative for each job title.

Yes, this will require significant work. The first year may be daunting. However, within a corporation of several thousand employees, dozens with the same job title, volunteers, especially those who realize a stake in the outcome, can be enlisted to help in this process. Today's employees are intelligent. They are so intelligent that, not only can they do their job, but they can also measure how good of a job that they and their peers are doing.

Average the Relative Weights

Once all performance ratings are submitted by the managers, Human Resources (HR) will perform a least squares fit (via Excel or computer program) of the five independent variables of each common measurement versus the one dependent variable of the overall rating to yield five coefficients (one for each common measurement) plus one y-intercept.

The purpose of this least squares fit is to reduce the bias of managers' individual weighting to arrive at a corporate weighting.

HR then uses these five coefficients and y-intercept to recalculate the overall rating of each employee.

Mathematically, it is possible that some coefficients could be negative, meaning that a higher rating of one measurement (e.g., initiative) will cause a negative effect on the overall rating. This outcome, though expected to be rare, does not fit into the rationale of performance evaluations. If so, reevaluate how to implement this averaging process, if at all.

Publish Historical Precedents

Courts within the United States follow a policy of *Stare Decisis*, which means, "to stand by decided matters." When the facts are substantially the same, current decisions are based previous decisions, called precedents.¹² Occasionally but rarely, the court may not follow a historical precedent, but does so with a well-written opinion explaining its rationale.

One of the problems with the current evaluation system is that we lose the historical precedents of how employees were evaluated in previous years. At the least, the employees being evaluated lose those precedents, or never had them in the first place. Management does not share what others did to succeed.

One solution is to publish the employee evaluations: their supporting evidence, their rationale, and their numerical rating for all seven measurements (five common measurements, one initial rating, plus one final rating). This solution will:

1. Demonstrate to the employees what it takes to succeed and
2. Build trust in leadership.

Immediately, the issue of confidentiality arises. Two solutions to the confidentiality issue are:

1. The employee signs a waiver to indemnify (hold harmless) the corporation, and/or
2. HR changes the names, projects, and other identifying information.

The key point is to provide a well-documented basis for the performance evaluation decisions.

Ask the Employees

Except for HR personnel and executive management, virtually no employee has been asked, “How shall we evaluate your performance?” before the performance period has begun. The process has always been that HR creates a new approach, then tells the employees.

By asking the employees for their input ahead of time you obtain their cooperation and, ideally, their embracing and advocating the new process.

How important are your employees, anyway?

References

-
- ¹ <http://www.raymondjames.com/careers/diversity.htm>, Raymond James Career Center, “Excellence Through Individuality”
 - ² <http://www.raymondjames.com/careers/education.htm>, Raymond James Career Center, “Education”
 - ³ <https://www.raymondjames.com/pr/130419.htm>, Raymond James Receives BISA Technology Innovation Award for Goal Planning & Monitoring Tool, April 19, 2013
 - ⁴ <http://www.raymondjames.com/advisorchoice/library/pr/140328.asp>, Raymond James Receives BISA Technology Innovation Award for Client Center Tool, March 28, 2014
 - ⁵ <http://www.raymondjames.com/about/mission.htm>, Mission Statement, Raymond James Financial, viewed June 2015
 - ⁶ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2917255>, Identifying and Avoiding Bias in Research, Christopher J. Pannucci, MD and Edwin G. Wilkins, MD MS, US National Library of Medicine, National Institutes of Health
 - ⁷ <http://www.statsref.com/HTML/bias.html>, Statistical Analysis Handbook, © 2015 Dr. M. J. de Smith
 - ⁸ http://www.storyofmathematics.com/19th_gauss.html, The Story of Mathematics, “19th Century Mathematics - Gauss
 - ⁹ <http://quotations.about.com/b/2010/10/08/how-fortunate-for-leaders-that-men-do-not-think.htm>, Adolf Hitler
 - ¹⁰ <http://www.triz-journal.com/triz-what-is-triz/> TRIZ – What is TRIZ? Katie Barry, Ellen Domb, Michael S. Slocum
 - ¹¹ <http://www.fastcompany.com/3041250/how-to-emulate-some-of-the-most-innovative-companies-out-there>, How To Emulate Some Of The Most Innovative Companies, Jose Resendiz, Fast Company, viewed May, 2015
 - ¹² Black’s Law Dictionary, Fourth Edition, Henry Campbell Black, West Publishing Company, St. Paul, 1968