

# PERSON-TO-PERSON RATING SCALE

Authored by  
**mohammad looti**

November 2, 2025

## RECOMMENDED CITATION

mohammad looti (2025). *PERSON-TO-PERSON RATING SCALE*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=62650>

## PERSON-TO-PERSON RATING SCALE

**Primary Disciplinary Field(s):** Industrial-Organizational Psychology; Psychometrics; Performance Appraisal

### 1. Core Definition

The Person-to-Person (PTP) Rating Scale is a specialized, comparative performance appraisal tool designed to standardize subjective evaluations by anchoring trait assessment not to numerical or purely verbal definitions, but to concrete human examples. This methodology contrasts the characteristics or performance traits of the individual being rated (the ratee) directly against the characteristics of a pre-established, representative reference group, often referred to as the comparison group or anchor group. The fundamental purpose of this structure is to reduce the cognitive burden and subjective interpretation often associated with abstract rating systems, forcing the rater to make a pragmatic, behavioral comparison.

In practice, the rating procedure requires the evaluator to identify the specific member within the anchor group whose displayed extent of the trait (e.g., initiative, technical skill, leadership) most closely matches that of the ratee. Once this closest match is determined, the ratee is formally assigned the score, rank, or designation previously attributed to that matching anchor individual. For example, if the comparison group consists of five individuals (rated 1 through 5, from low to high initiative), and the ratee is deemed to possess the same level of initiative as the comparison group member designated '4,' the ratee receives a '4' rating. The PTP scale thus transforms an abstract judgment into a structured, comparative matching exercise.

This rating method belongs to a broader family of comparative appraisal techniques, yet it distinguishes itself by relying entirely on the human element as the standard unit of measurement. While traditional **Likert scales** use descriptive phrases or numerical points as anchors, the PTP scale uses actual or conceptualized employees who represent discrete points on the performance continuum. This reliance on tangible comparison points is intended to mitigate common rating errors such as the **halo effect** or central tendency bias, as the rater is compelled to assess variation relative to known standards rather than relying on personal internal standards of 'average' or 'excellent.'

### 2. Etymology and Historical Development

The origins of the Person-to-Person rating scale method are primarily situated within the early development of psychological testing and personnel selection during the first half of the 20th century. A critical need arose, particularly in military and large industrial organizations, to evaluate complex, subjective traits--like officer potential or emotional stability--where precise, objective metrics were unavailable. Early rating systems were frequently criticized for being highly

susceptible to rater idiosyncrasies, leading to inconsistent and unreliable evaluations.

The PTP concept, sometimes referenced under the umbrella of 'man-to-man rating scales' in older literature (reflecting the terminology of the era), was an early psychometric attempt to overcome these deficiencies through standardization. By providing raters with concrete behavioral references--the comparison group--the systems sought to homogenize the definition of performance traits across different evaluators. These early scales proved instrumental in large-scale assessment programs during World War I and World War II, where rapid and reliable selection and promotion decisions were paramount, paving the way for the sophisticated performance appraisal methods used in modern **Industrial-Organizational Psychology**.

Although modern performance appraisal systems frequently favor methods like Behaviorally Anchored Rating Scales (BARS) or Graphic Rating Scales for their ease of administration and explicit behavioral definitions, the foundational principle of the PTP scale--the use of standardized human benchmarks for comparative assessment--remains deeply influential. Contemporary variations of comparative assessment, such as **forced distribution** and paired comparison techniques, owe a debt to the PTP method's commitment to relative rather than absolute measurement, acknowledging the inherent difficulty humans face in making objective, abstract judgments about performance.

### 3. Operational Mechanics and Implementation

Implementing a successful Person-to-Person rating scale involves rigorous preparatory work, focusing on the careful selection and calibration of the anchor group. The process generally begins by identifying the critical performance dimension (or trait) to be evaluated, such as leadership effectiveness or conflict resolution ability. Once defined, organizational experts, often including high-level managers and HR professionals, must collaboratively select a small group of individuals who collectively represent the entire spectrum of performance for that trait, ranging from the absolute lowest acceptable level to the highest achievable level within the organization. This selection ensures that every possible performance level has a tangible, human reference point.

Following the selection of the anchor group, each member must be carefully studied and consensus must be reached on the specific numerical value or rank they represent regarding the defined trait. This calibration step is perhaps the most crucial, as the assigned value must accurately reflect their demonstrated behaviors. Rater training is then essential, ensuring that evaluators understand the history and performance profile of each anchor individual and know how to use these profiles as the fixed, objective standard against which all other employees will be judged. Raters are specifically instructed not to use their personal opinion of the anchor individuals, but rather the established organizational profile of their performance relative to the trait.

During the actual evaluation phase, the rater compares the ratee not against a written job

description, but directly against the memory or description of the anchor individuals. The rater must continually ask, "On this trait (e.g., strategic thinking), is the ratee better than Anchor C but not as effective as Anchor D?" This mental comparison continues until a precise match is located. This forced-choice mechanism, requiring a direct match to a defined standard person, is designed to reduce the tendency of raters to cluster scores in the middle (central tendency error) or rate everyone highly (leniency error). The output is a comparative score that benchmarks the ratee against the known human standard, providing a relative measure of performance.

#### 4. Key Characteristics

**Concrete Anchoring:** The most defining characteristic is the use of **human models** (or composites of human performance) as the explicit standards for comparison, rather than abstract adjectives (poor, average, excellent) or numerical assignments. This tangibility is intended to stabilize the rating process.

**Forced Comparative Judgment:** PTP scales necessitate a relative evaluation. Raters are not asked how "good" a performance is in isolation but where the performance falls relative to the established set of human benchmarks. This forces discrimination between performance levels.

**Trait-Specific Focus:** The scale is typically constructed to evaluate specific traits (e.g., dependability, resourcefulness) rather than broad, holistic job performance. A separate anchor group might be required for each trait evaluated, increasing the administrative overhead but improving precision for each dimension.

**High Developmental Costs, Low Usage Costs:** The initial investment in selecting, validating, and training raters on the anchor group is substantial. However, once established, the scale itself is relatively quick to administer, as the raters merely select the best match rather than composing detailed commentary or interpreting complex behavioral descriptions.

#### 5. Significance and Impact

The Person-to-Person rating scale holds historical and practical significance primarily because it provided an early, structured methodology for standardizing subjective assessment, laying the groundwork for subsequent behavioral rating systems. Its greatest impact lies in demonstrating that by framing performance evaluation as a choice among concrete, fixed alternatives, organizations could achieve higher inter-rater reliability than was possible with simple graphic scales. This increased reliability is particularly valuable in environments where the criteria for success are difficult to quantify, such as evaluating potential for rapid promotion or assessing intangible leadership qualities.

Furthermore, the PTP method influenced the design of modern talent management and human

resource systems that rely on peer comparison. The concept of benchmarking employees against internal high-performers, which is central to succession planning and high-potential identification programs, directly echoes the PTP philosophy. By forcing organizations to identify and codify the traits exhibited by exemplary employees (the anchors), the PTP process inadvertently helps define and reinforce organizational culture and performance expectations, providing clear, if comparative, models for development.

However, the PTP method's most enduring legacy might be its successful challenge to purely absolute rating systems. It highlighted that human raters are inherently relative judges, and that reliable measurement often requires incorporating this relativity into the scale design itself. While the direct implementation of the PTP scale has diminished due to ethical and logistical concerns (discussed below), its comparative philosophy is ubiquitous in contemporary performance management, influencing everything from forced ranking systems to the statistical normalization of performance scores to ensure fairness across large groups.

## 6. Methodological Challenges and Ethical Debates

Despite its conceptual strengths in reducing rater biases, the Person-to-Person rating scale presents significant practical and ethical hurdles that have contributed to its decreased use. Logistically, the initial requirement of identifying and validating the anchor group is demanding. If the anchor individuals leave the organization or if their performance changes, the entire scale's foundation is compromised, necessitating time-consuming re-validation. Furthermore, ensuring that the chosen anchors truly represent universally recognized performance levels across all departments or functional areas is often impossible, leading to a potential lack of perceived **face validity** among employees being rated.

Ethically, the PTP scale introduces potential issues concerning privacy and fairness. When performance is explicitly tied to a specific person--even if the anchor individuals are anonymized--there is an inherent risk that employees may deduce or discover the identities of the benchmarks. This can lead to resentment, political maneuvering, or the perception of unfair judgment if an employee feels they are being negatively compared to a specific, potentially controversial, individual. Legally, basing crucial employment decisions (like promotion or termination) on comparisons to named individuals, rather than against explicit, behaviorally defined job requirements, can complicate defense against discrimination claims if the rating system appears subjective or biased toward the characteristics of the chosen anchor group.

Finally, the PTP scale often provides limited developmental feedback. While an employee receives a numerical rating, the accompanying feedback is inherently comparative ("You matched the performance of Anchor 3"), which fails to provide the specific, actionable behavioral details necessary for improvement. Unlike BARS, which links scores directly to critical behavioral

incidents, PTP scales require the rater to provide supplemental narrative feedback to make the score meaningful for employee growth, adding complexity to the rating process that the scale was originally designed to simplify.

## 7. Relationship to Modern Comparative Methods

The PTP scale serves as a conceptual antecedent to several modern comparative evaluation techniques used in organizational psychology. It shares the goal of standardizing subjective assessment with the **paired comparison method**, where every employee is compared to every other employee in the rating pool. However, PTP is significantly more efficient because it only requires comparison against a small, fixed set of anchors (usually 5 to 9 people), whereas paired comparison results in a rapidly growing number of comparisons ( $N * (N-1) / 2$ ).

Furthermore, PTP principles inform the design of modern forced distribution systems, such as the 'rank and yank' method, which mandate that a fixed percentage of employees must be placed into predefined performance categories (e.g., 20% top performers, 70% average, 10% poor). While forced distribution uses statistical requirements to achieve relative ranking, the PTP scale uses concrete human standards to achieve the same relative assessment. Both methods recognize that absolute rating scales often suffer from inflation and lack differentiation, necessitating a mechanism that forces raters to spread out scores and make difficult, comparative judgments.

Ultimately, the Person-to-Person rating scale represents a critical phase in the evolution of performance appraisal: the transition from purely impressionistic judgments to structured, behaviorally informed comparative assessment. While rarely used in its original form today, the focus on human benchmarks continues to influence how organizations conceptualize, measure, and manage talent relative to internal standards of excellence.

### Further Reading

[Performance Appraisal \(Wikipedia\)](#)

[Psychometrics \(Wikipedia\)](#)

[Industrial and Organizational Psychology \(Wikipedia\)](#)