

# Practical Exploration of Subgroup Differences in Performance Ratings

Kevin G. Smith, M.A.

Tiffany M. Bennett, Ph.D.

July 19, 2011

# Overview

- Previous Research
- Practical Implications
- Disconnect
- Techniques to Assess
- Conclusion

# Previous Research

- Accurate measurement of job performance has stimulated much interest from researchers, practitioners, and managers alike (Murphy, 2008)
  - The breadth of factors and cyclical research findings remain a challenge even for well-informed practitioners in implementing effective performance management
- Persistent subgroup differences have been observed when comparing ratings made on Black and White employee's performance across many studies (McDaniel & McKay, 2006)
  - Understanding subgroup differences is important from several standpoints (e.g., ethical, moral, legal, fairness)
  - It is unclear why subgroup differences exist, especially within specific organizational contexts

# Previous Research

- Drawing on some social categorization and social identity theories (Tajfel & Turner, 1979), rating bias has been one potential explanation
  - Disagreement regarding the need for further research on this topic, its relative prevalence, and significance (Stauffer & Buckley, 2003)
- The gold standard (Pulakos, White, Oppler, & Borman, 1989) for evaluating bias in performance ratings requires a within-subjects design whereby multiple raters rate the same ratee
  - In practice, this method is virtually impossible to utilize
- When within-subjects designs are not feasible, research that controls for ability, education, and experience suggests small main effect Black-White differences (Waldman & Avolio, 1991)

# Practical Implications

- Subgroup differences in performance ratings are common and should not come as a surprise to a practitioner
- Expected magnitude of subgroup differences is hard to anticipate without considering several important moderators (McKay & McDaniel, 2005)
  - Measurement method (objective versus subjective)
  - Measurement level (single item versus scale)
  - Criterion type (task/contextual performance, absenteeism, etc.)
  - Use of the rating/data source (e.g., published research, dissertations, administrative purposes)
  - Demography of the organization (low % minority versus highly-diverse)
- The root cause of subgroup differences in performance ratings has **not** been unequivocally proven

# Practical Implications

- The research on subgroup differences combined with current legal requirements in the United States **deserves** a high level of organizational interest
  - Disadvantage is legally defined as adverse impact
  - Members of a protected group receive disproportionate outcomes compared to the majority
  - Typically measured by 80% rule or statistical significance test; for example:
    - 75% of males and 53% of females receive highest rating
    - $53\%/75\% = 71\%$  (violation of rule)
- If an organization's practices show adverse impact, the organization is vulnerable to legal challenges
- However, performance ratings and associated payouts (bonuses, salary) are often not categorical in nature

## Disconnect between Implications and Amount of Exploration

- The practical implications of subgroup differences are not uniformly recognized by those who are in charge of implementing performance management
- When faced with subgroup differences, some stakeholders rely on one hunch or another (some plausible, others not) without systematic investigation
- Unless there are obvious signs otherwise, many assume that rating differences are due to performance differences rather than any combination of other plausible hypotheses
- Monitoring means see if it happens again next year
- Race/ethnicity is a sensitive topic that some avoid addressing head-on
- Evaluation methods are underutilized and rarely integrated (just statistical analysis or focus groups)

# Scalable but Integrated Evaluation Techniques

## 1) Statistical Analysis

- Baseline analyses
- Advanced analyses
  - Move on to evaluation technique 2 if conclusions are not definitive

## 2) Performance Evaluation Reviews

- Review of the actual information supervisors used to document employee performance
  - Move on to evaluation technique 3 if conclusions are not definitive

## 3) Surveys and Interviews

- Address the employees' and other stakeholders' perceptions of fairness

## 4) Integrated Report

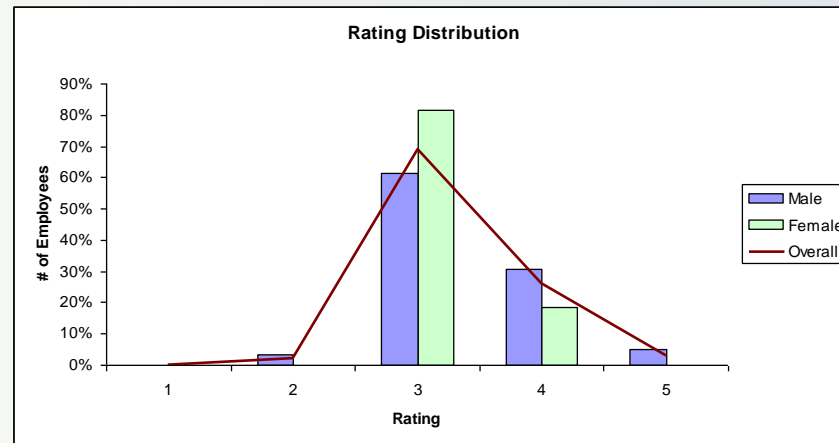
# Baseline Statistical Analyses

- Conduct baseline statistical analyses to determine the extent to which subgroup differences are present in the operational performance data
  - If small or non-existent ( $d=0.0-0.20$ ) additional analysis is not necessary
  - If moderate or large, additional evaluation techniques are required

Guidance for Interpreting Subgroup Differences Effect Sizes	
$d \leq .20$	Small and unlikely to produce significant differences in outcomes (i.e., payouts, bonuses) across subgroups
$.20 < d \leq .50$	Moderate and may produce different outcomes (i.e., payouts, bonuses) across subgroups; monitor carefully
$.50 < d \leq .75$	Needs attention; consult with expert for further analysis
$d > .75$	Large and will likely produce different outcomes across subgroups (i.e., payouts, bonuses); consult with expert

# Supplementary Analyses

- Focus on identifying where in the organizational strata (occupation, band level, locations) differences are most pronounced
- Conduct multiple regression analyses including relevant factors in research (e.g., experience, ability, education)
- If race effects are still significant after controlling for relevant variables, additional analysis is necessary



# Performance Evaluation Reviews

- In the Federal Government evaluation reviews are already accomplished through the completion of the Office of Personnel Management’s Performance Appraisal Assessment Tool
  - Current set of criteria answer important general system health questions but do not adequately address issues related to subgroup differences
- Additional steps can be taken to review potential bias in written narratives that incorporate contemporary perspectives on discrimination (Dipboye & Colella, 2005) as discussed in Wilson, 2010
  - Collect targeted sample of performance evaluations

Level of Performance						
	Low Performers		Average Performers		Good Performers	
Black	Negative mention	Positive mention	Negative mention	Positive mention	Negative mention	Positive mention
White	Negative mention	Positive mention	Negative mention	Positive mention	Negative mention	Positive mention

# Performance Evaluation Reviews

- Trained and calibrated I/O Psychologists rate the information contained on the performance evaluations according to structured rating criteria
- After analyzing the data, this method answers the following questions related to subgroup differences:
  - Are there more negative mentions of task factors in supervisor's summaries across races/ethnicity within the same level of performance?
  - Are supervisors more likely to mention positive knowledge and learning in justifying performance across races/ethnicities?
  - Are supervisors more likely to mention positive interpersonal aspects of contextual performance across races/ethnicities?
  - Are supervisors more likely to mention positive effort, initiative, or enthusiasm across races/ethnicities?
  - Do supervisors offer more negative mentions at lower levels of performance?
  - Do supervisors offer more positive mentions at higher levels of performance?
  - Do supervisors emphasize different performance factors at different levels of performance?

# Performance Evaluation Reviews

- After analyzing the data, this method answers the following questions related to general system health:
  - Are the performance evaluations prepared according to performance management policies?
  - Do the performance evaluations indicate any systems-related training needs?
  - Did the performance ratings receive adequate narrative justification?
  - Are average and poor performers getting actionable feedback?

# Survey and Interviews

- Surveys and focus groups to address important system stakeholders' (employees, managers, practitioners) perceptions
  - Perceptions and organizational climate with regards to performance management
  - Perceptions regarding fairness in the distribution of developmental activities, high-visibility projects, and richness of professional networks, etc.
  - Analyze data to determine themes

# Conclusion

- In practice, subgroup differences in performance ratings are not considered as often as they should be
- Comparing qualitative and quantitative data cross multiple methods gives the broadest understanding behind what may be driving subgroup differences within a specific organization
- Scalable, integrated evaluation techniques are relatively low effort and can be completed by a small team of I-O Psychologists within a 4-6 week period using data that is at the organization's disposal
  - Surveys and focus groups may take longer due to increased coordination time
- Results inform performance management system improvements, adverse impact mitigation strategies, and diversity recruitment strategies

# Contact Information

- Kevin G. Smith, M.A.
  - 703-812-5340
  - Kevin.Smith@pdri.com
- Tiffany M. Bennett, Ph.D.
  - 703-812-3047
  - Tiffany.Bennett@pdri.com