Best Practices in Assessment Centers

Reducing “Group Differences” to a Phrase for the Past

Michael D. Blair
CWH Management Solutions
mblair@cwhms.com

Overview

- General Introduction to Assessment Centers.
- Subgroup Differences in Assessment Centers.
- CWH Assessment Center Results
- Research and Best Practices to Reduce Subgroup Differences

Purpose of Assessment Centers

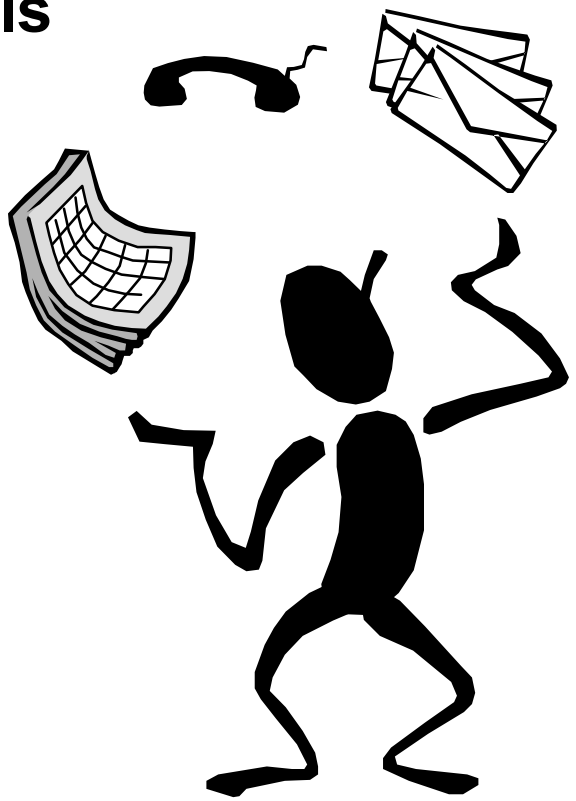
- Assessment Centers are used for a wide variety of purposes:
  - Selection
  - Placement
  - Promotion
  - Identification of Management Potential
  - Training
  - Career Development

Assessment Center Validity

- Research has consistently demonstrated that Assessment Centers successfully predict a variety of important outcomes.
  - Job Performance
  - Management Potential
  - Training Performance
  - Career Development

Types of Dimensions Assessed

- Organizational and Technical Skills
- Management and Practical Skills
- Tactical Skills
- Interpersonal Skills
- Leadership Skills
- Communication and Presentation Skills
- Written Communication Skills



Common Exercises

- Written Exercise
- Structured Interview
- Group Discussion
- Oral Presentation
- In-Basket
- Role-Play
  (Subordinate Conference)
- Emergency Scenario
  (or Tactical Exercise)

Subgroup Differences in Assessment Centers

- Research on the subgroup differences in Assessment Centers has been mixed.
  - Some studies have found no differences.
  - Other studies have found significant differences between White candidates and Black candidates.
- Generally accepted in the field that Black-White subgroup differences exist in most measures, including assessment centers.

Typical Subgroup Differences Across a Variety of Commonly Used Measures

Measure	Score Difference (in SD-Difference Units)
Cognitive Ability	1.00
Personality (The Big Five)	-0.04 to 0.21
Structured Interview	0.23
Biodata	0.33
Video Situational Judgment	0.43
Paper Situational Judgment	0.61
Assessment Center	0.20 to 0.60 (0.40)

Partially adapted from Ployhart & Tsacoumis (2001).

Typical Subgroup Differences in Common Assessment Center Exercises

Exercise	Score Difference (in SD-Difference Units)
In-Basket	0.35
Subordinate Meeting (Role Play)	0.03
Group Discussion	0.25
Project Presentation	0.27
Project Discussion	0.39
Team Preparation	0.40
Overall Score	0.40

From Goldstein, Yusko, Braverman, Smith, & Chung (1998).

Typical Subgroup Differences In Common Police Assessment Center Exercises

Exercise	Score Difference (in SD-Difference Units)
In-Basket	1.15
Subordinate Counseling (Role Play)	0.14
Spoken Incident	0.31
Briefing & Training Subordinates	0.26
Overall Score	0.62

From Goldstein Ruminson, Yusko, & Smith, (2001).

Typical Subgroup Differences In Common Police and Fire Assessment Center Exercises

	In-Basket Score Difference (in SD-Difference Units)	**Role-Play and Incident Command Score Difference** (in SD-Difference Units)
Fire	0.56	-0.12
Police	0.67	0.25

Fire collapsed across Lieutenant, Captain, & Battalion Chief
Police collapsed across Sergeant, Lieutenant, Captain

From SHL Landy-Jacobs (unpublished data)

Subgroup Differences From Recent CWH Fire and Police Assessment Centers

Exercise	Fire Score Differences (in SD-Difference Units)	Police Score Differences (in SD-Difference Units)
In-Basket (Oral)	-0.20	-0.24
Emergency Scenario	-0.02	-0.20
Role-Play	-0.18	0.20
Structured Interview		-0.15
Oral Presentation	0.42*	-0.06
Written Exercise		0.48
Overall Score	-0.16	-0.14

CWH Data collected from 2000-2003 and collapsed across Fire Driver/Engineer, Lieutenant, & Captain (n = 477) and Police Sergeant & Lieutenant (n = 77)

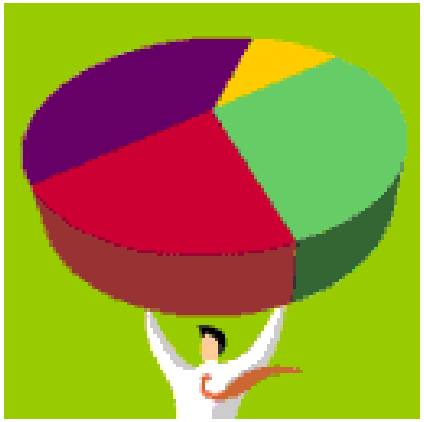
*Fire oral presentation based on n = 35)

Research and Best Practices to Reduce Subgroup Differences

- Focus on the entire process, not just a part of the process
  - **Job Analysis & Test Plan**
  - **Exercise Choice & Development**
  - **Test Administration**
  - **Assessor Training and the Rating Process**
  - **Candidate Feedback**

Job Analysis

- Reducing group differences begins with the job analysis.
  - Typical job analysis overemphasizes cognitive ability.
  - Increase the job analysis domain to capture the full range of KSAs.
  - Focus on non-cognitive elements.
  - Job analysis is so heavily weighted toward cognitive aspects, that non-cognitive aspects get buried.

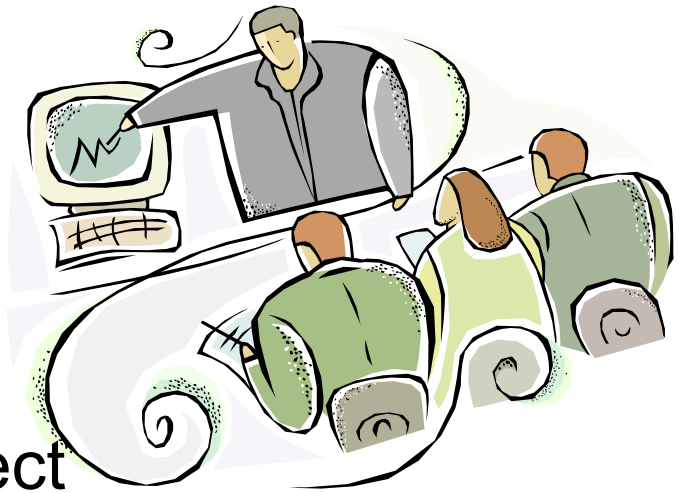


Test Plan

- Use a rational approach
  - Keep it simple and use what you know about the job.
  - May reduce AI over complex mathematical models.
  - May increase validity.
- Have SMEs provide weighting data.
  - Review and weight individual exercises.
  - Reduces group differences by reducing overemphasis on cognitive loaded exercises.
- Use a unit weighting approach
  - If can't use SMEs, equal weight the exercises.
  - Results in lower group differences.

Exercise Choice & Development

- Cognitively loaded exercises lead to adverse impact.
  - The higher the cognitive load, the greater the B-W score difference.
  - ACs overemphasize cognitive aspects of exercises.



- Interactive exercises better reflect most jobs and have lower adverse impact.
- Response mode should vary and reflect the job
  - Why is an in-basket a written exercise?
  - CWH uses oral in-baskets – compare the difference!

Test Administration

- Reduce the information processing and reading comprehension requirements in the candidate materials.
  - Should accurately reflect the job.
- Allow ample time for preparation and ample face time with the assessors.
  - Minorities perform poorer on speeded tests.
  - More exercise time = more opportunity for observation and more interaction.

- Conduct candidate preparation in-between exercises, not all at once.
  - Logistically more difficult, but more realistic and better results.
- Use diverse assessor panels
  - Increase validity and decrease subgroup differences.

Assessor Training

- Bridge behavioral observation (traditional approach) with frame-of-reference (FOR) training.
  - FOR increases rating accuracy, reliability, and validity.
  - May result in fewer recorded behavioral observations.

- Training model:
  - Focus on dimensions and behaviors related to each dimension.
  - Define behaviors along the continuum within each dimension.
  - Link behaviors to dimensions “on-the-fly”, not after the exercise.
  - Focus on recording behavioral observations for feedback.
  - Conduct practice sessions using “live” mock candidates.
  - Feedback to the assessors regarding rating accuracy and FOR.
  - Debrief practice exercises.

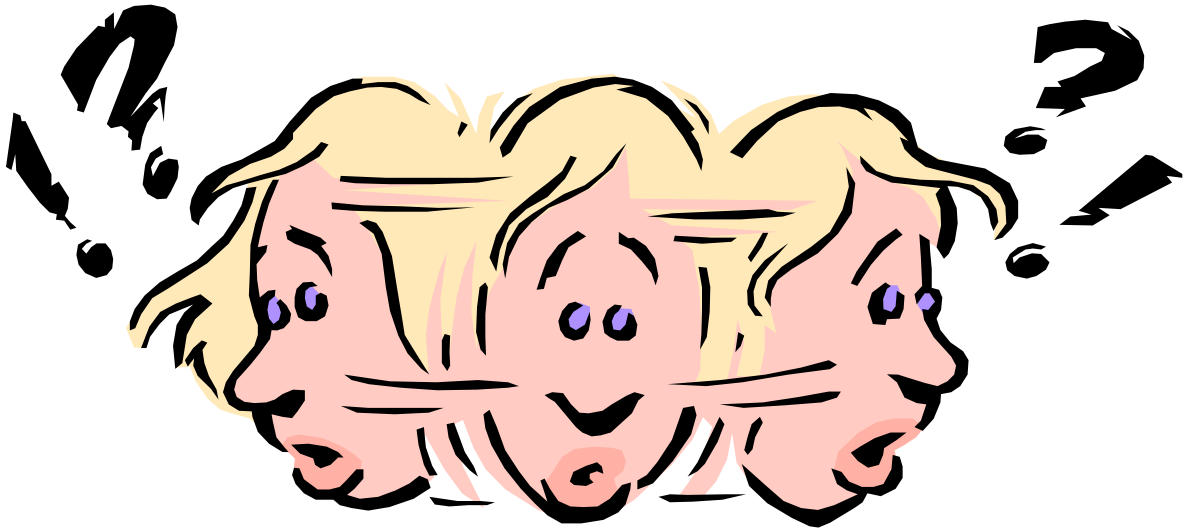
The Rating Process

- Use fewer dimensions
  - Just like memory, it's 7 +/- 2.
  - Increases the number of behaviors correctly assigned.
  - Reduces rating errors due to bias.
- Use behavioral checklists that link example behaviors to the dimensions.
  - Increases accuracy, reliability, and validity of ratings.
  - Reduces burnout due to cognitive load.
  - Allows more time to observe actual behavior.
- Do not force consensus, but insist on assessor discussions after each candidate.
  - Discuss ratings and observed behavior.
  - Pooling of comments & sharing of observations.
  - Keeps assessors are on same page and in focus.

Candidate Feedback

- **Provide candidate feedback**
  - Anecdotal accounts indicate that candidate feedback can reduce subgroup score differences over time.
- **Feedback should:**
  - Identify strengths, weaknesses, and suggestions for improvement.
  - Include direct statements from assessors.
  - Include roll-up reports so candidates can compare performance to the group.

QUESTIONS



References

Carliss, S.A. & Allwood, V.E. (1997). Managerial assessment centers: What is being rated? *Australian Psychologist, 32, 101-105.*

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72, 493-511.*

Gaugler, B.B. & Thornton, G.C. (1989). Number of dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology, 74, 611-618.*

Goldstein, H.W. (2000). *Selection instruments: Which predict best?* Keynote Address at the 24th Annual IPMAAC Conference on Personnel Assessment, Washington, DC.

Goldstein, H.W., Ruminson, K.C., Yusko, K.P., & Smith, D.B. (2003). The impact of test composite score weighting procedures on Black-White subgroup differences. Unpublished manuscript.

Goldstein, H.W., Yusko, K.P., Braverman, E.P., Smith, D.B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology, 51, 357-374.*

Hough, L.M., Oswald, F.L., Ployhart, R.E. (2002). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: issues, evidence, and lessons learned. *International Journal of Selection and Assessment.*

Hornick, C.W. & Axton, T.R. (1998). Weighing issues: Balancing low adverse impact and high validity. Paper presented at the 106th Annual Convention of the American Psychological Association, San Francisco, CA.

References (Cont)

LeBreton, J.M., Gniatczyk, L.A., & Migetz, D.Z. (1999). *The relationship between behavior checklist ratings and judgmental ratings in an operational assessment center: An application of structural equation modeling.* Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology.

Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology, 86, 255-264.*

Lievens, F. & Klimoski, R.J. (2001). Understanding the assessment centre process: Where are we now? In, Cooper, C.L. & Robertson, I.T. (Eds.), *International Review of Industrial and Organizational Psychology, Volume 16.* John Wiley & Sons, Ltd.

Ployhart, R.E. & Tsacoumis, S. (2001). *Strategies for reducing adverse impact.* Paper presented at the February, 2001, Workshop of the Personnel Testing Council of Metropolitan Washington, D.C. (PTC/MW).

Raymark, P.H., Scmit, M.J., & Guion, R.M. (1997). Identifying potentially useful personality constructs for employee selection. *Personnel Psychology, 50, 723-736.*

Reilly, R.R., Henry, S., Smither, J.W. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. *Personnel Psychology, 43, 71-84.*

Schleicher, D.J., Day, D.V., Mayes, B.T., & Riggio, R.E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology, 87, 735-746.*