

CWH Research, Inc.

The First Shall Be Last and the Last Shall Be First: Does Presentation Order and Scheduling Impact Candidate Scores in Assessment Centers?

Thomas J. Carnahan, Ph.D

Chris W. Hornick, Ph.D

John M. Ford, Ph.D



1

Overview

Where the Data Came From

Presentation Order Impact

Adverse Impact of Assessment Centers

Discussion



2

What is included in the dataset?

- Police and Fire Promotional Assessment Centers
- Three Criteria to be an Assessment Center
 - Multiple exercises
 - Multiple trained observers
 - Measure job related behaviors
- January 2006 through April 2008



3

Dataset Descriptors

- 48 Assessment Centers Included
- Total N of 1075 Candidates
 - Race: White = 752; Black = 115; Hispanic = 55; Native American = 16
 - Gender: Male = 891; Female = 54
- Broken Down by AC Exercise
 - N = 3251
 - Tactical = 151
 - Hot Seat = 511
 - In-basket = 420
 - Oral Presentation = 456
 - Role Play = 1,015
 - Structured Interview = 537
 - Standardized within exercise
 - Converted to a Mean of 80 with a SD of 5



4

What was investigated?

- Order within the Assessment Center
- Presentation order in front of the trained raters
- Day within a multi-day Assessment Center
- Morning versus Afternoon
- Presenting immediately after lunch
- Race and Gender (Dean, Roth and Bobko, 2008)



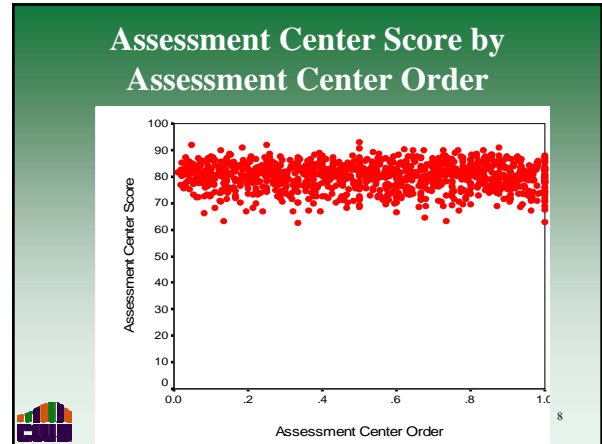
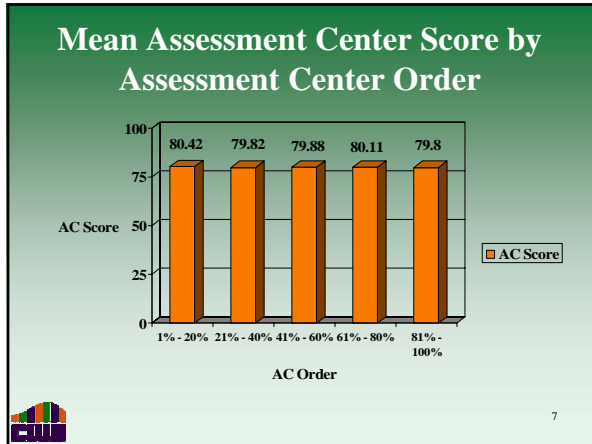
5

Order in the Assessment Center

- Does going first, last, or in the middle of an Assessment Center affect scores?
- Methodology
 - Created a variable that divided the candidate's order number by the total number of candidates in the Assessment Center
 - Correlated this new variable with the scores on the Total Assessment Center
- $r = -.04$

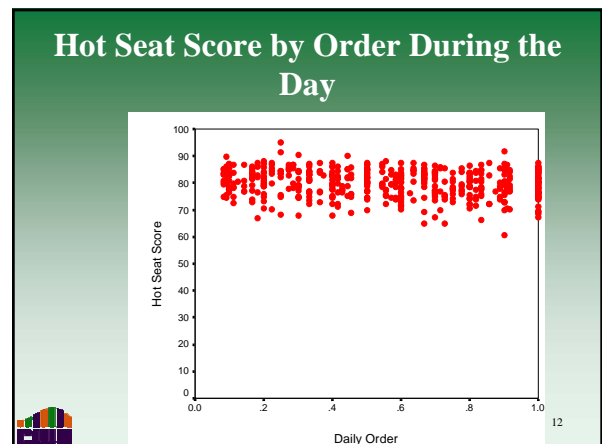
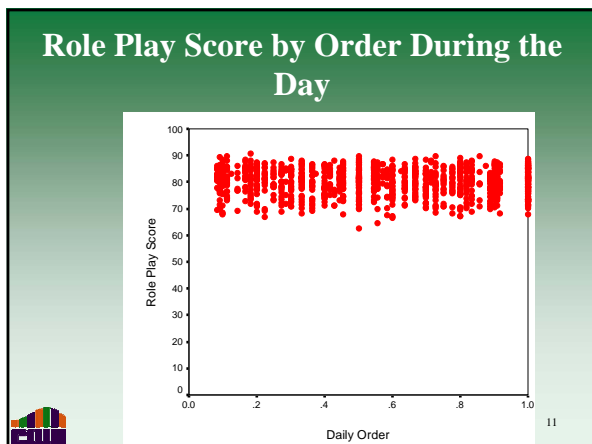
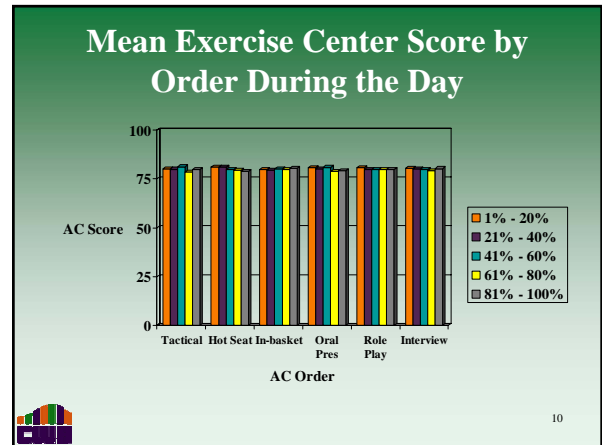


6



Exercise Presentation Order in front of the Trained Raters

- Does going first, last, or in the middle of the day affect scores?
- Methodology
 - Created a variable that divided the candidate's order in front of the assessors (BY EXERCISE) by the total number of candidates assessed for the day
 - Correlated this new variable to their Assessment Center EXERCISE Score
- Correlations range from .05 to -.18.
 - Only one correlation was "significant" (hot seat)
 - Practical differences were very small



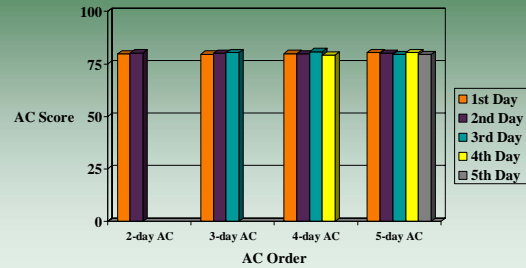
Day within a Multi-Day Assessment Center

- Does the day that you went matter in a multi-day assessment center?
 - 2-day AC: 203 candidates
 - 3-day AC: 204 candidates
 - 4-day AC: 76 candidates
 - 5-day AC: 338 candidates
- Methodology
 - ANOVA
 - Independent Variable: Day Assessed
 - Dependent Variable: Total Assessment Center Score
- Results
 - 2-day AC: $F = 0.47, NS$
 - 3-day AC: $F = 0.36, NS$
 - 4-day AC: $F = 0.29, NS$
 - 5-day AC: $F = 0.56, NS$



13

Mean Assessment Center Score by Day Assessed



14

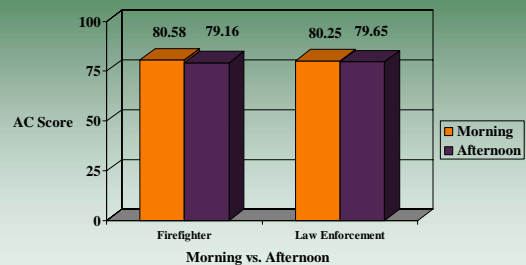
Morning versus Afternoon

- Does whether or not the candidate presents in the morning versus the afternoon affect their Assessment Center score?
 - Firefighters: 316 Morning candidates, 218 afternoon candidates
 - Law Enforcement: 315 Morning candidates, 226 afternoon candidates
- Methodology
 - Created a variable that identified whether the candidate performed before lunch or after lunch
 - Correlated this new variable with the scores on the Total Assessment Center
- Results
 - Firefighters: $F = 10.68, p < .01$
 - Law Enforcement: $F = 1.85, NS$



15

Mean Assessment Center Score by Morning vs. Afternoon



16

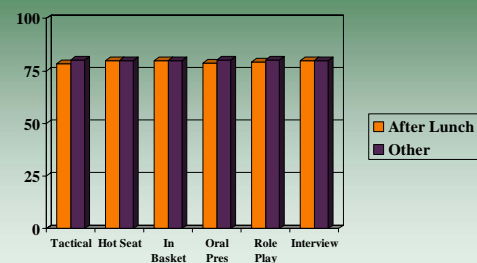
Presenting Immediately After Lunch

- Does whether or not the candidate is the first to present after lunch affect their Assessment Center Exercise score?
- Methodology
 - Created a variable that identified whether the candidate was first after lunch (score of 1) or not (score of 0)
 - Correlated this new variable to their Assessment Center EXERCISE Score
- Correlations range from .00 to -.09.
 - None were significant
 - Practical differences were very small



17

Mean Exercise Score by First After Lunch



18

Race and Gender Differences

- We wondered...What else could the data could tell us?
- We investigated all of the above again by race and gender to determine if there were differences – there were none
- What about race and gender differences in general?



19

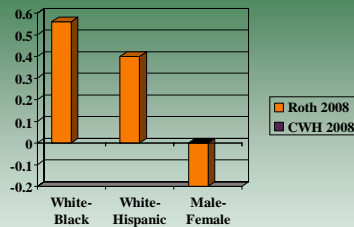
Dean, Roth, and Bobko (2008)

- First meta-analysis to ever ask the question: What is the SD difference for race and gender for Assessment Centers?
- *d* statistic represents the difference in group means divided by the pooled standard deviation of the relevant groups.
 - Example: a *d* of 0.5 would indicate that the groups differ by ½ of a standard deviation
- Their dataset contained 10 Assessment Centers for Black-White candidate comparisons, 8 Assessment Centers for White-Hispanic comparisons, and 9 Assessment Centers for Male-Female comparisons
- Again, CWH used the dataset created from 48 Assessment Centers from across the US (San Diego, CA to New Haven, CT)



20

Dean, Roth, and Bobko Results

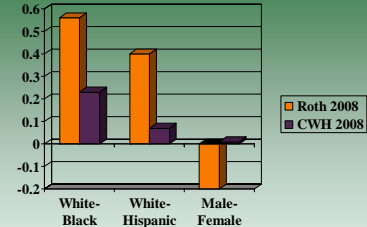


- They concluded that Assessment Centers are not as free from Adverse Impact as researchers think (that is not below a *d* of .30)



21

CWH (2008) Results



- Assessment Centers are still a viable method by reducing impact against protected groups and providing excellent content validity



22

Placing These Results in Context

- Results Reflect ONLY our Data and Assessment Centers
- General Assessor Training
 - Department specific training
 - Rating errors and Rater bias
 - Scoring system
 - Mock candidates
 - Mentoring from consultants



23

Placing These Results in Context

- Focus of Assessor Training
 - Rate candidate based on observable behaviors
 - It is what the client department wants, not the rater's department
 - Reinforcement of roles and expectations
 - Overall Score which allows for comparisons
 - Assessors usually only see one exercise which reduces influence from other exercises
 - Process for dealing with panel score differences



24



Future Directions

- Are there other differences between professions?
- Video vs Live presentation formats
- Officer vs Command Level candidates
- What if we took the data back to 1996? What trends would we see?
- Rater ethnicity differences and similarity/dissimilarity to candidates
- Experience differences of candidates in assessment centers



25

Discussion

Any Questions???



26