

# Techniques for Improving the Fairness of Large-Sample Assessments Requiring Multiple Rater Panels

**Matt Gruver**  
Senior Manager, CPS

**Jack Clancy**  
Director, Clancy & Associates



# Program Objectives

- Discuss strategies for reducing the effects of rater inconsistencies across rater groups
- Examine logistics of setting up and conducting structured interviews and other procedures for assessing large candidate groups

# Adding structure increases reliability and validity

- Research has demonstrated the usefulness of structured interviews for selection
- Meta analysis by Schmidt and Hunter reported overall validity of structured interview of .51 vs. .37 for unstructured interviews

# Adding structure increases reliability and validity

- Schmidt and Hunter's research also found that structured interviews provide additive validity to cognitive, written tests
- Research on structured interviews was summarized in a literature review by Campion, Palmer, and Campion

**Campion paper lists  
15 components of  
interview structure:  
seven pertaining to  
interview content and  
eight that addressed  
evaluation**



# Interview content components

1. Base questions on a job analysis
2. Ask exact same questions of each candidate
3. Limit prompting, follow-up questioning, and elaboration on questions
4. Use better types of questions

# Interview content components

5. Use longer interview or larger number of questions
6. Withhold or control ancillary information
7. Do not allow questions from candidate until after the interview

# Interview evaluation components

8. Rate each answer or use multiple scales
9. Use detailed anchored rating scales
10. Take detailed notes
11. Use multiple interviewers

# Interview evaluation components

12. Use same interviewers across all candidates
13. Do not discuss candidates or answers between interviews
14. Provide extensive interviewer training
15. Use statistical rather than clinical prediction



# Large-sample assessments pose special problems

- Components 2 and 12 are difficult for large candidate groups
  - Ask exact same questions
  - Use same interviewers across all candidates
- Large candidate groups typically require multiple forms and multiple interview panels



## Common Solution

- We typically fix the problem afterwards by standardizing candidate scores within rater groups
- Corrects for assumed differences in candidate scores as a function of panel differences caused by rater bias (overly conservative, overly lenient)

Panel 1
25
30
35
40
45
50
55
60

Panel 2
30
35
40
45
50
55
60
65

Panel 1	Panel 1 Z	Panel 1 Stand		Panel 2	Panel 2 Z	Panel 2 Stand
25	-1.42887	55.71		30	-1.42887	55.71
30	-1.02062	59.79		35	-1.02062	59.79
35	-0.61237	63.88		40	-0.61237	63.88
40	-0.20412	67.96		45	-0.20412	67.96
45	0.20412	72.04		50	0.20412	72.04
50	0.61237	76.12		55	0.61237	76.12
55	1.02062	80.21		60	1.02062	80.21
60	1.42887	84.29		65	1.42887	84.29

**Problem – you don't know  
when differences between  
panels are a result of rater  
bias or chance  
knowledge/ability differences  
between candidate groups**

Question – what happens  
when score differences result  
from real differences between  
candidate groups?

Panel 1
12
14
15
12
14
16
18
20

Panel 2
16
18
19
22
22
24
24
25

Panel 1	Z-Score	Stand Score		Panel 2	Z-Score	Stand Score
12	-1.1161	58.84		16	-1.6202	53.80
14	-0.4018	65.98		18	-1.003	59.97
15	-0.0446	69.55		19	-0.6944	63.06
12	-1.1161	58.84		22	0.23146	72.31
14	-0.4018	65.98		22	0.23146	72.31
16	0.31251	73.13		24	0.84867	78.49
18	1.02683	80.27		24	0.84867	78.49
20	1.74115	87.41		25	1.15728	81.57

# Practical Solutions

- Practice test
- Focus on training
- Rotate panelists

# Practice Test

- SME judgments
- Mock interviews and exercises

# Rater Training

Should focus on appropriate interviewing/assessment techniques and the system to be used:

1. Behavioral observation, classification, and evaluation
2. The testing instrument

# Panel Rotation

- Stratify panels
- Rotate panel members  
(half-day or full-day)
- Rater consensus

# Rater Schedule

	Panel A	Panel B	Panel C	Break
Monday PM	Lt. White	Lt. Black	Lt. Green	Sgt. Johnson
	Sgt. Smith	Sgt. Jones	Sgt. Wilson	
Tuesday AM	Lt. White	Lt. Black	Lt. Green	Sgt. Wilson
	Sgt. Johnson	Sgt. Smith	Sgt. Jones	
Tuesday PM	Lt. White	Lt. Black	Lt. Green	Sgt. Jones
	Sgt. Wilson	Sgt. Johnson	Sgt. Smith	
Wednesday AM	Lt. White	Lt. Black	Lt. Green	Sgt. Smith
	Sgt. Jones	Sgt. Wilson	Sgt. Johnson	

# Candidate Schedule

	Panel A	Panel B	Panel C
8:00 – 8:40 am	1	2	3
8:40 – 9:20 am	4	5	6
9:20 – 10:00 am	7	8	9
10:00 – 10:40 am	10	11	12
10:40 – 11:20 am	13	14	15
11:20 – 12:00 pm	16	17	18
LUNCH			
1:00 – 1:40 pm	19	20	21
1:40 – 2:20 pm	22	23	24
2:20 – 3:00 pm	25	26	27
3:00 – 3:40 pm	28	29	30
3:40 – 4:20 pm	31	32	33
4:20 – 5:00 pm	34	35	36

# Practical Solutions Revisited

- Structure
- Training
- Panel make-up
- Consensus
  - Absolute
  - Within 1-point
- Stay involved