

# **To Band or Not to Band: Is that the Question?**



**IPAC 2011**

1

**FRANK IGOU, LOUISIANA TECH UNIVERSITY**

**JEFF FEUQUAY, PSYCHOLOGY-LAW CENTER**

**MICHAEL BLAIR, SPRINT**

**CHRIS HORNICK, CWH RESEARCH, INC.**

**JOHN FORD, CWH RESEARCH, INC.**

# Some Purposes of this Panel

2

- ❖ Continue the discussion with people who are “seasoned assessment and selection veterans”
- ❖ Introduce or further clarify the topic to people who are early career
- ❖ Hopefully, share some experiences, thoughts, ideas or insights on when it’s best to band or not to band.

# Panel Members

3

- ❖ Frank Igou, Louisiana Tech University
- ❖ Jeff Feuquay, Psychology-Law Center
- ❖ Michael Blair, Sprint
- ❖ Chris Hornick, CWH Research, Inc.
- ❖ John Ford, CWH Research, Inc.

Unfortunately, Drs. Chris Hornick and John Ford are working with clients. We will present their content, but they are not able to join us today.

# Brief Overview of Banding

4

- ❖ The traditional top-down model of selection is one in which the scores obtained from valid selection test are said to bear a linear relationship to measures of job performance.
- ❖ Higher-scoring applicants are expected to demonstrate higher levels of measured job performance than lower-scoring applicants.

# Less than Perfect Prediction

5

- ❖ However, it is well documented that selection tests do not predict job performance perfectly (Murphy & Myers, 1995; Murphy, 1994).
- ❖ The relationship between selection tests and criterion measures of job performance is typically found to range from  $r = 0.20$  to  $r = 0.50$  (Campion, Outtz, Zedeck, Schmidt, Kehoe, Murphy and Guion, 2001).

# Errors in Prediction

6

- ❖ Because of errors in prediction, the classical selection model is probabilistic rather than deterministic.
- ❖ For example, selection tests cannot predict whether an individual job candidate with a selection test score or civil service score of 94 will demonstrate better job performance than another applicant with a score of 91.

# Prediction

7

The general prediction is that hiring groups of higher-scoring job candidates will lead to higher levels of job performance on average...

when examining individual's job performance and looking at aggregate job performance for member of that job group.

# Individual Errors in Prediction

8

Individual errors in prediction create situations where a job candidate may score high on a valid selection test and is selected, but demonstrates a less than adequate level of job performance, while another job candidate may score somewhat lower on the same valid selection test, is not selected, but is capable of a high level of job performance.



# False Hits and False Misses

- ❖ “False hits” or “erroneous acceptances” can lead to occasions in which individuals who lack sufficient qualifications for a job are hired.
- ❖ “False misses” or “erroneous rejections” can lead to not hiring qualified individuals and further may preclude the hiring of truly qualified individuals who may be members of an EEO protected class.

# Reported Demographic Differences

10

Some research has reported that the mean test scores of some protected groups

(e.g., African-Americans and Hispanics; women on standardized mathematics tests)

are about one standard deviation below the mean test score of referent groups.

(e.g., Whites or males; McKinney & Collins, 1991, Gottfredson, 1986).

# Demographic Differences

11

**There is still disagreement over the cause of these reported differences.**

- ❖ Some maintain that there are actually “real world” differences among groups in innate intelligence, developmental opportunities, etc...
- ❖ Some believe the testing processes are biased. For example, different groups use language differently.

# Errors in Prediction

12

False misses or erroneous rejections due to error in prediction may reduce employment opportunities for minority group members and can perpetuate the effects of past discrimination on job candidates from lower scoring minority groups.

(Murphy, 1994; Hartigan & Wigdor 1989).

# Linear Scores or Groups of Scores?

Sproule (1984) discussed the use of grouping similar scores:

“On written tests for example, one can use the standard error of measurement or a multiple thereof, as a basis for establishing ranges or confidence intervals of scores, from the passing score up or from the highest score down, to form groups...

... All raw scores within a given interval would then be considered the same final score.” (p.390)

# Using Error of Measurement for Creating Bands

14

Cascio, Zeddeck, Outtz and Goldstein (1994)

suggest the use of the **standard error of difference** (henceforth “SED”) as that the proper statistic for determining whether two scores are reliably different.

In its simplest computational formula, the *SED* is simply the product of the *SEM* and the  $\sqrt{2}$

Scores should be placed groups or bands.

The size of these score band may be calculated as:

$$\text{Band Width} = C \times \text{sd}_x \times (1 - r_{xx})^{1/2} \times 1.414$$

# Bandwidth

15

In the previous slide,  $r_{xx}$  refers to the reliability of the test, and the term “C” refers to the normal deviate that corresponds to the desired level of confidence.

For example, a C value of 1.96 corresponds to a 95% confidence interval. Thus, if one wanted to establish bands that were 95% confidence intervals, one would set the bandwidth at approximately 2 *SEDs*.

# What is Typically Believed

16

- ❖ Top-down selection tend to produce the most adverse impact but yield the highest utility (i.e., value-added return in employee performance).
- ❖ Banding methods that use minority preference, especially sliding bands, are least likely to produce adverse impact and the most likely to help increase workforce diversity. Further, the utility loss produced by such methods will be marginal, that is no greater than .20 - .30 of a standard deviation.



# Critics of Banding

17

Critics of banding maintain there is no justification for the use of these methods.

For example, Schmidt and Hunter (1994) maintain that it isn't sufficient to examine a single selection event. The real utility is found over multiple selection events and that top-down methods will in the aggregate sense, return maximum utility.

# Utility and Unit of Analysis

18

Cascio, Zeddeck, Outtz & Goldstein (1994) suggested that the ultimate utility of any selection system cannot always be examined at the individual job-candidate or employee level.

# Another Way to View Test Utility

19

Picture a situation in a diverse community in which the police officers are selected from the top of a list of eligible candidates. If the “top-of-the-listers” are white, you might have the highest performers, but officers may not have full acceptance by the community in which they work, and consequently the job may be more difficult to perform.

# Litigation Support Costs

- ❖ Another source of utility may be money saved by not having to spend it on litigation.
- ❖ The Uniform Guidelines on Employee Selection mention that assessment and selection professionals need to consider alternative selection procedures which may be equal in prediction but have less adverse impact.
- ❖ It is possible that the use of banding could fulfill the same end.

# Published Research - 1

To date, there appear to be only two published researched articles.

1. Cascio, Zeddeck, Outtz & Goldstein (1991; 1994), which used a very diverse large sample (N= 6503).

This study compared the outcomes from strict top-down selection, top-down within group selection and four banding method.

# Cascio et al. Discussion

22

However, most recruitments produce smaller lists of eligible candidates. Typical lists of eligible candidates may contain 300 or fewer candidates.

Schmitt and Hunter (1994) referred to this data set as an “anomalous data set” because of its size, demographic composition and lack of mean differences between white and minority candidates.

# Published Research - 2

2. Sackett and Roth (1991; 1994), used a Fortran-based Monte Carlo study to examine the external validity of Cascio et al.

This study used parameters of the data in the Cascio et al. study and examined hypothetical outcome from banding under different selection rates, reliability estimates, and bandwidths.

Note: This study was a simulation; it is possible that the distributions of real selection data may deviate somewhat from assumptions of normality.

# Literature Search

- ❖ There are numerous book chapters (e.g, Cascio and Aguinis) and theoretical articles (e.g. Murphy, 1994).
- ❖ A search of EBSCO and PSYCHINFO (databases which list published research) produced no additional articles as of July 7, 2011.
- ❖ There are no articles which examine typical outcomes of the use of banding in federal, state or city/county agencies, or typical outcomes from the use of banding in large private organizations such as health care or telecommunication organizations.



# To Band or Not to Band?

25

How does an assessment and selection professional decide when to band or not to band?

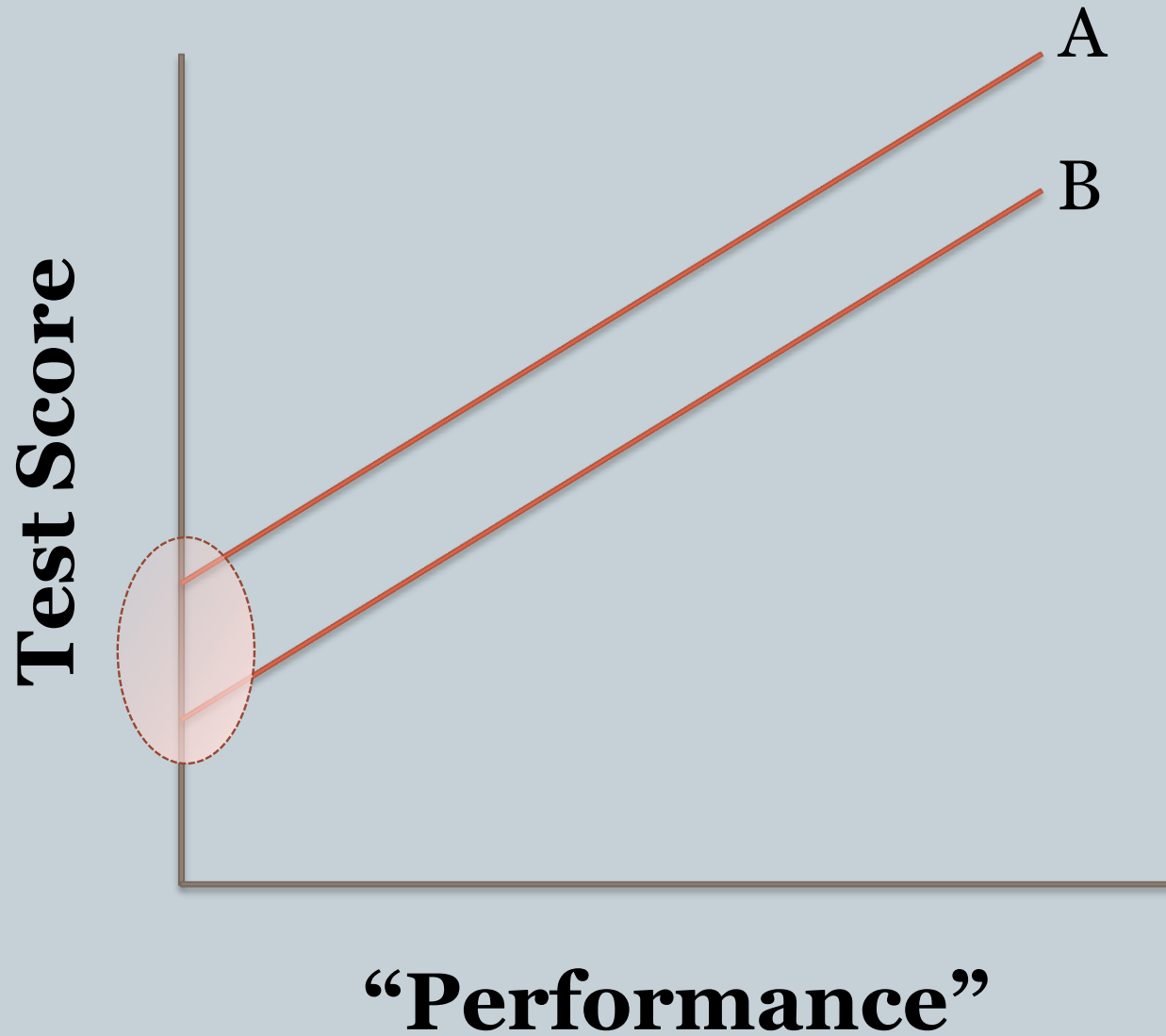
The panel today will share experiences, thought and insights that hopefully will be useful to those who are considering changing their way their organization uses scores from selection tests.

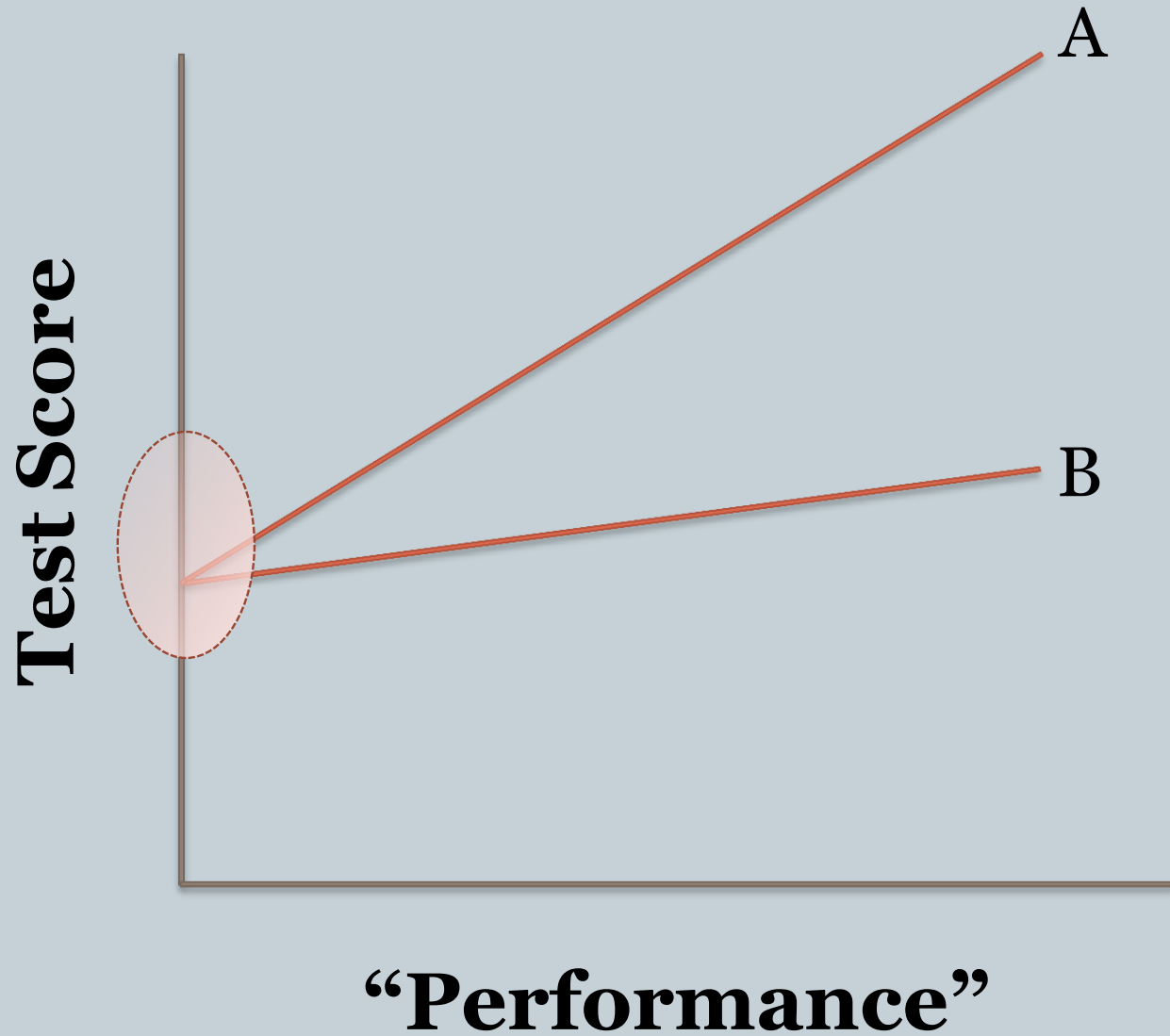
# Which displeases courts more?

26

❖ **disparate impact**

❖ **disparate treatment**





**So, the question is:**

29

**WHY BAND SCORES?**

# What's the motive for banding scores?

30

- Had the City reviewed the exam results and then adopted banding to make the minority test scores appear higher, it would have violated Title VII's prohibition of adjusting test results on the basis of race.

Ricci v. DeStefano, 129 S.Ct. 2658, 2683 (2009)

# What's the motive for banding scores?

31

- “We have no doubt that if banding were adopted in order to make lower black scores seem higher, it would indeed be a form of race norming, and therefore forbidden.”
- “But it is not race norming per se. In fact it's a universal and normally an unquestioned method of simplifying scoring by eliminating meaningless gradations.”

Judge Posner

Chicago Firefighters Local 2 v. Chicago, 249 F. 3d 649, 655-56 (CA7 2001)

So, the *new* question is:

32

**What are the legal reasons  
for banding scores?**



# **Banding in the Public and Private Sector**



**MICHAEL BLAIR**  
**SPRINT**

# A Very Brief History of Banding

34

- More recently banding has been associated with statistical bands or normative bands
  - Everyone within a group (band) is considered to be equally qualified according to the selection criteria used
- But...banding has been used for as long as people have been making hiring decisions
  - Rule of X
  - Highly recommend, recommend, not recommend
  - Top 1/3, middle 1/3, bottom 1/3
  - Sorting resumes into piles
    - ✦ Yes/no/maybe...interview/get more info/reject...hire/no hire
  - Like it or not...you are banding!

# Why Banding May Make Sense

35

- Measurement error in testing
  - The test score varies from the true score
- Tests only measure part of the criterion space
  - Banding allows for expansion of the criterion space at a later stage in the process
- Flexibility to hiring manager
  - Allows for final selection/placement to take into account additional factors such as person-team or person-position fit
- Can reduce adverse impact
  - Most effective if using minority preference within a band and using sliding bands as selection occurs

# Establishing Bands

36

- **Administrative**
  - Typically drawn for convenience
- **Statistical**
  - Based on the standard error of measurement (SEM)
  - Calculated using the standard error of the difference (SED), test reliability, and confidence interval
  - Debates on the appropriateness of statistical bands and the proper calculation have been ongoing for 20+ years
- **Normative**
  - Based on very large scale sample sizes – thousands or 10's of thousands
  - Data is normed and bands are based on percentile scores
- **Criterion or performance based**
  - Bands established based on performance indicators
  - Used in assessment scoring (BARS, BSS), but not used to determine bands based on the results of test scores

# Fixed Versus Sliding Bands

37

- Fixed bands
  - Bands are set and do not change based on who is selected from the band
  - Everyone in a band is selected/rejected before moving to the next band
- Sliding bands
  - Bands change based on who is selected from the band
  - When an individual is selected, the band is recalculated before making additional selections
    - ✦ Often “recalculation” is done by simply moving the next highest score up into the band

# Banding in the Public Sector

38

- Administrative bands
  - Qualified, not qualified
  - Highly qualified, qualified, less qualified, not qualified
  - Scores from 100-90, 89-80, 79-70, 69-60
  - First 20%, second 20%, third 20%, etc.
- Rule of 'X' – variation on administrative bands
  - Rank order of scores is used to generate list based on a rule of 'X'
  - Rules of 3, 5, and 10 are common
  - Used with and without replacement (i.e., sliding and fixed bands)
- Statistical bands
  - Traditional is 1 SED and 95% confidence interval
  - Broader bands use 1.5 SEDs or increase confidence interval to 99%
  - Occasionally see bands based on 2 SEDs
- Bands are typically established after testing is complete
  - But based on the rule or policy
- Validation studies are rarely done to help establish the bands or determine effectiveness after implementation
  - Validation is limited to establishing the job-relatedness of the assessment

# Banding in the Private Sector

39

- Administrative bands
  - Interview – Yes, no, maybe
  - Meets basic qualifications, does not meet basic qualifications
  - Does not meet requirements, meets requirements, meets preferred requirements
- Normative bands
  - Typically see top 30%, middle 40%, bottom 30%
    - ✦ Green, yellow, red
    - ✦ Highly recommended, recommended, not recommended
  - Also see variants based on business needs
    - ✦ Top versus bottom half
    - ✦ Top 20%, next 30%, bottom 50%
- Statistical bands
  - Not often seen in the private sector
- Bands are typically established before testing begins
- Validation studies to establish and revise (“optimize”) the bands are the norm, especially for the Fortune 500
  - Also used to provide ROI and establish the business case for assessments

# Reasons for Banding: Public vs Private

40

- **Public Sector**
  - It mostly comes down to legal liability...
    - ✦ Eliminate or mitigate adverse impact
    - ✦ Correction for past transgressions
  - Sometimes driven by desire to increase diversity or provide a means to handle candidate volume
- **Private Sector**
  - Primary driver is business outcome and/or ROI
  - Secondary driver is often candidate management
  - Adverse impact or diversity is rarely a factor
    - ✦ Addressed during test development & validation, as well as through recruitment/sourcing

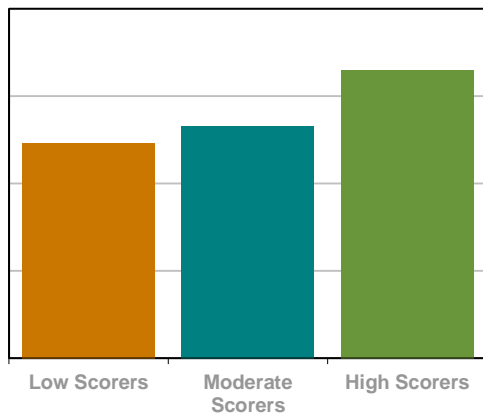


# Banding and Business Outcome Front-Line Care Improvements

41

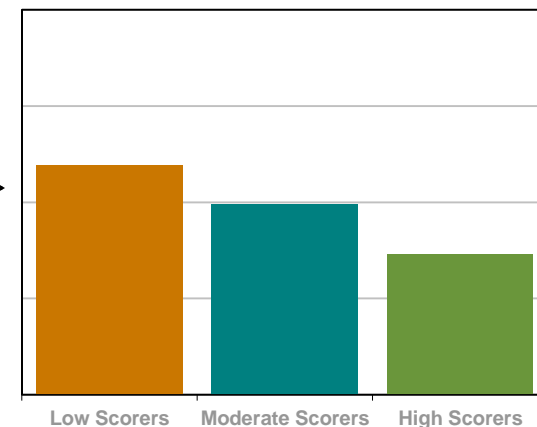


## Issue Resolve



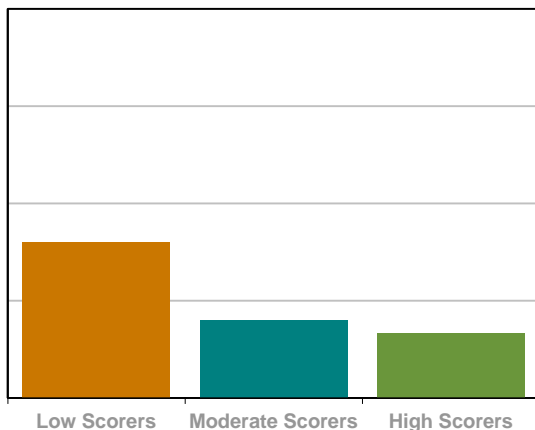
← High scorers resolve 5% more issues than low scorers

## Average Handle Time



High scorers handle calls 46 seconds faster than low scorers →

## Transfers

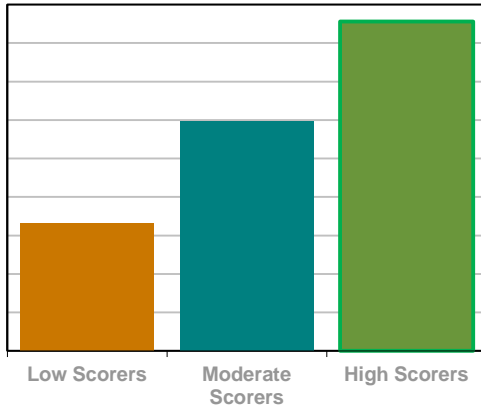


← High scorers transfer 5% fewer calls than low scorers

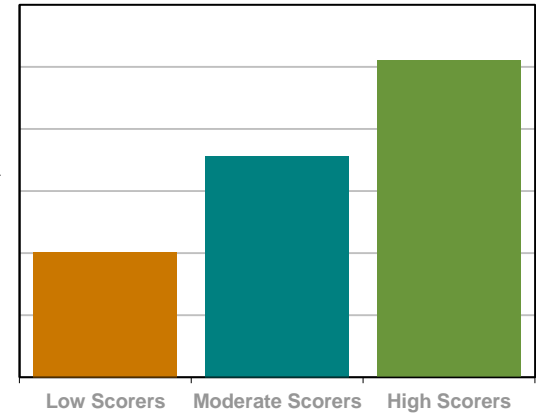
# Banding and ROI

## Front-Line Sales Improvements

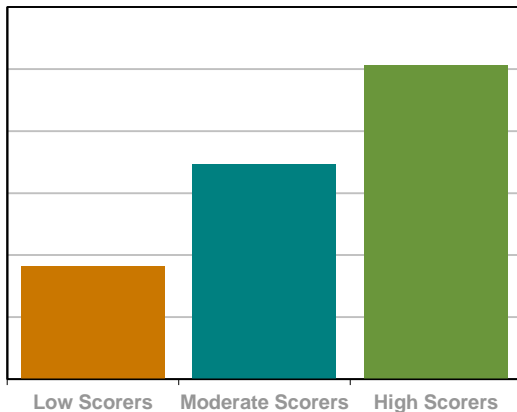
42



← Monthly Accessory Revenue  
Greens +\$263, +16%  
compared to Reds



→ Monthly Net Activations  
Greens +1.54, +11%  
compared to Reds



← Monthly Upgrades  
Greens +3.24, +8%  
compared to Reds

# Potential Pitfalls of Banding

43

- Not a panacea for diversity or adverse impact
  - Minority preference can have a significant effect
    - ✦ Recent court cases indicate that minority preference may not be legal
  - Banding can actually work against diversity if the test has no AI
- Loss of information and lower utility
  - Particularly for broad bands
    - ✦ Is a 94 really the same as an 91...maybe...
    - ✦ Is a 94 really the same as a 71...probably not...
- “Pass the buck” from HR to the hiring manager
  - Especially if the bands are large
- Can be misused or abused by the hiring manager
  - Selection from within the band based on poor or no criteria
  - Selection from within the band based on the wrong criteria

# Selecting From Within the Band

44

- Still under legal regulations and professional guidelines
  - Must be able to defend the process
- Need to help the hiring manager
  - What does he/she need?
  - What did the test measure?
  - How can the process be structured, yet efficient?
- Establish a Procedure and Follow It
  - Random selection
  - Targeted selection on critical components
  - Stacked rank selection
  - Other

# Banding is not a Magic Bullet

45

- Banding will not correct poor recruitment practices
  - Need diverse and qualified applicant pools
- Banding will not correct poor, biased, or discriminatory selection tests
  - Use non-biased, reliable, and valid selection measures prior to banding
    - ✦ It is possible to create tests with little or no AI
    - ✦ Low AI leads to diversity in the score distribution and increases diversity in the bands

# Banding References

46

- Reference list is broken out into several categories
  - Essential reading
  - Statistical articles
  - Legal perspectives
  - Additional sources

# **Practical Alternatives to Statistical Banding**



**CHRIS W. HORNICK**  
**JOHN M. FORD**

# Alternate Techniques for Establishing Bands

48

- Performance criteria or level of expected performance (i.e., criterion-referenced bands)
- Equal opportunity considerations
- Ease of administration
- Cost of administration
- Pre-established standards
- Professional judgment



# Reasons to Avoid Statistical Banding

49

- There are many theoretical and psychometric arguments against Statistical Banding.
  - There is not a consensus among psychometricians and testing professionals regarding the science of banding.
  - Frank Schmidt, Jerard Kehoe, and others have argued that the assumptions supporting banding are fundamentally flawed and that statistical banding is logically inconsistent.
- There are also several practical reasons for establishing bands using the alternate techniques discussed on the previous slide rather than the using statistical bands.

# Counterproductive Outcomes Often Result

50

- Under typical test conditions ( $r = .80$ , normal distribution), the top band can include 38% of score range and 25% of candidate pool (Schmidt in Campion et al., 2001).
- Statistical bands are frequently equal to or larger than a standard deviation (Kehoe & Tenopyr, 1994).
- Assuming typical levels of reliability and validity, Kehoe & Tenopyr demonstrated the following:
  - Candidates who score at the top of a statistical band are 25 times more likely to outscore candidates who score at the bottom of the same band when taking the test again.
  - Candidates who score at the top of a statistical band are twice as likely to outperform candidates who score at the bottom of the same band.
  - Candidates who score at the top of a statistical band can be twice as likely to meet successful performance standards as compared to candidates who score at the bottom of the same statistical band.

# Candidate Understanding and Trust

51

- Candidates generally don't understand statistical banding.
- When candidates don't understand the selection process, they become distrusting and suspicious.
- Distrust in the process has many negative consequences:
  - More appeals and challenges
  - More lawsuits
  - Less willingness to accept other parts of the process
  - Less participation in future processes
  - Poorer relationships with HR and organizational leadership
  - Bad press and/or poorer relationships with the general public

# Statistical Banding Is Often Somewhat Disingenuous

52

- Statistical banding is often used in a somewhat disingenuous manner:
  - The stated reason for using statistical banding is that small differences in test scores are not meaningful— “We are just practicing good science.”
  - However, the real reason for using statistical banding is often that it allows organizations to insert non-score related criteria into the process of making selection systems under the cover of science.
- This allows organizations to avoid having serious discussions about their values and articulating these values to their stakeholders and formally incorporating them into their organizational processes.
- The more serious risk is that organizations will be charged with using inappropriate criteria or using criteria in an arbitrary or subjective manner. Even if untrue, the appearance that the organization tried to cover up what they were really doing can give the organization a black eye or bad name.

# Adverse Impact and Diversity

53

- The courts have generally supported the use of statistical banding if selection within bands is not based on minority preference. However, selecting based on minority preference has not been supported.
- Research has consistently demonstrated that the use of statistical banding without minority preference does not reduce adverse impact or increase diversity (Barrett, Doverspike, and Arthur, 1995 ; Gutman and Christiansen, 1997; Sackett and Roth, 1991; Sackett and Wilk. 1994).
- However, banding would likely reduce adverse impact if pass points and bands were established based on the likely distribution of minority candidates in the test distribution.
  - For example, if a specific test score has resulted in acceptable adverse impact ratios and diversity in previous administrations over thousands of candidates, this score could be set as the bottom of the top band (i.e., everyone who achieves this score would move on to the next step in the process). This technique would be more likely to help organizations meet their diversity goals than statistical banding.

# Real World Example—Background

54

- Large law enforcement agency in Midwest.
- For their promotional processes, they previously used statistical banding to group candidates into score bands.
  - When promotional opportunities occur, the Chief is given an alphabetized list of all of the candidates in the top band.
  - The Chief is allowed to promote anyone within the band.
  - No one from the next band can be promoted until everyone in the previous band has been promoted.
- There was an extremely high level of mistrust by candidates in how the bands were established.
  - Candidates did not understand why bands were used, how they were calculated, why they were so large, or why there were different numbers of candidates in each band (even though the previous consultant had explained the process to them several times).
  - During orientations, candidates were extremely hostile toward the banding process and openly accused individuals of manipulating the bands .

# Real World Example—Process Change

55

- We recommended that the department change the way that bands are established.
- For the assessment center, we created a scale to use when assessing candidates.

Rating	Description
0 - 49	Totally unacceptable, lacking any effort
50 - 59	Needs significant improvement
60 - 69	Approaches minimally acceptable
70 - 79	Meets or slightly exceeds minimum requirements
80 - 89	Very good, approaching excellent performance
90 - 99	Excellent to outstanding performance

- We worked with subject matter experts to develop specific benchmarks related to this scale.
- We spent two full days of training and practice with the assessors to help them master the scale and benchmarks.

# Real World Example—Results

56

- We established bands based on the scale that was used in the assessment center.
  - For example, candidates with scores over 90 were placed in the “Excellent Band.”
- We (the consulting company) sent a document with the banding results directly to every candidate. We also provided an explanation for how the bands were established to every candidate.
- There were no challenges or appeals to the process based on the banding process.
- During the orientations for the next process, there was no controversy or resistance when we announced that the bands would be established in a similar manner. Candidate acceptance had jumped considerably.



# Final Thoughts - 1

57

- If an agency is going to move away from selecting based strictly on the rank order of test scores, they should establish bands and processes that are simple, understandable, easy to explain to candidates, and consistent with explicit organizational values (such as efficiency, budget consciousness, equal opportunity, social values, etc.).
- If you wouldn't feel comfortable explaining your process, your true purpose for using bands, the way bands are established, and the criteria used to select within the bands to your candidates, the public, the media, and a court of law, you should re-think your selection process.

# Final Thoughts - 2

58

- Using more practical techniques to establish bands will generally provide more flexibility in meeting organizational goals, better adverse impact and diversity results (assuming that you use this as one of the criteria for establishing your bands), and more positive candidate reactions (as well as fewer complaints, appeals, challenges, etc.).