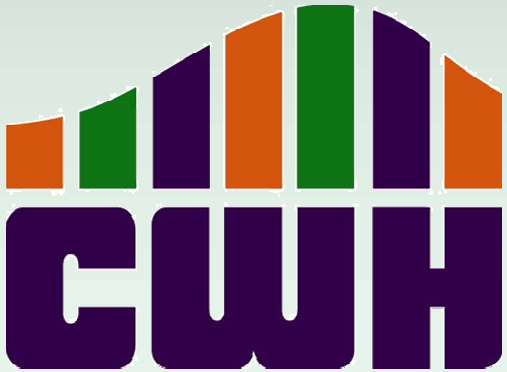
CWH Research, Inc.

Moving Beyond "Eeny, Meeny, Miny, Moe": What Factors Should Guide the Evaluation of Selection Tests

John M. Ford

CWH Research, Inc.



Overview

- **Critical Assumptions**
- **7 Important Considerations When Evaluating Selection Tests**

Critical Assumptions

- **Tests have value.**
- **All tests are not the same.**
- **Tests are part of an overall organizational environment—Changing to a new test will not magically change your organization by itself.**

#1: Don’t let the tail wag the dog—Take control of your process

- **The RFP process is not conducive to making informed decisions.**
  - **Don’t let test providers decide what information you should consider.**
- **Evaluating tests requires professional judgment.**
  - **You must ask the right questions and evaluate the evidence.**

#1: Don’t let the tail wag the dog—Take control of your process

- **Don’t forget future and hidden costs.**
  - **Inefficient performance**
  - **Increased training/remedial training/retraining**
  - **Lawsuits**
  - **Turnover**
  - **Grievances**
  - **Disciplinary problems**
  - **Accidents**

#2: There is no such thing as a valid test

"Validity refers to the degree to which evidence and theory support *the interpretations of test scores* entailed by *proposed uses* of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a *sound scientific basis for the proposed score interpretation of test scores required by proposed uses* that are evaluated, <u>not the test itself</u>. When test scores are used or interpreted in more than one way, each intended interpretation must be validated" (*Standards for Educational and Psychological Testing*, 1999; p. 9).

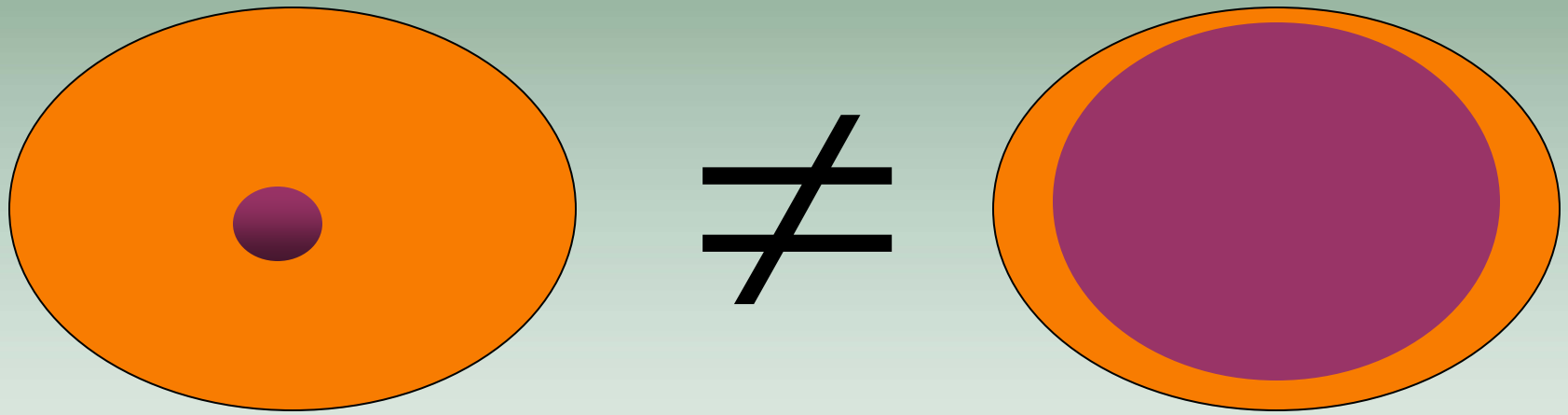
#3—Not all validation evidence is equal

- **Validity should not be treated as a categorical variable in your decision-making.**
- **Validation evidence should be evaluated along a continuum.**
- **This guideline applies to evidence regarding content relevance (i.e., content validity).**

Example: Not All Content Relevance Evidence is Equal.

Job Domain

Test Domain



Test 1 **Test 2**

Corollary—Adverse impact is also not a continuous variable

- **Adverse impact should also be evaluated on a continuum.**
  - **Although they both violate the 4/5ths rule, an AI ratio of .70 is preferable to .20.**
  - **Similarly, 1.00 is preferable to .80.**
- **Higher AI ratios provide a variety of results:**
  - **More diversity in your organization**
  - **Greater likelihood of meeting the 4/5ths rule in individual samples**
  - **Lower likelihood of grievances, EEOC investigations, lawsuits, and bad press**

#4: Context matters!!!

- **Validity cannot be properly evaluated without knowledge of the validation process.**
  - **Get the technical report.**
  - **Validation study circumstances should match your circumstances.**
- **Every validation study should include a job analysis or analysis of work.**
  - **Is the job domain appropriately defined?**
  - **Are the job requirements similar to your position?— This is necessary to transport validation evidence.**
  - **Are the test components defined in a manner consistent with the job domain?**

#4: Context matters!!!

- **Use of test should match your process/needs**
- **Validity coefficients are not an island—they provide very little information without context.**
  - **Is the sample appropriate for your agency?**
  - **Is the criterion related to important aspects of the job (and your job)?**
  - **Is the validity coefficient corrected or uncorrected?**

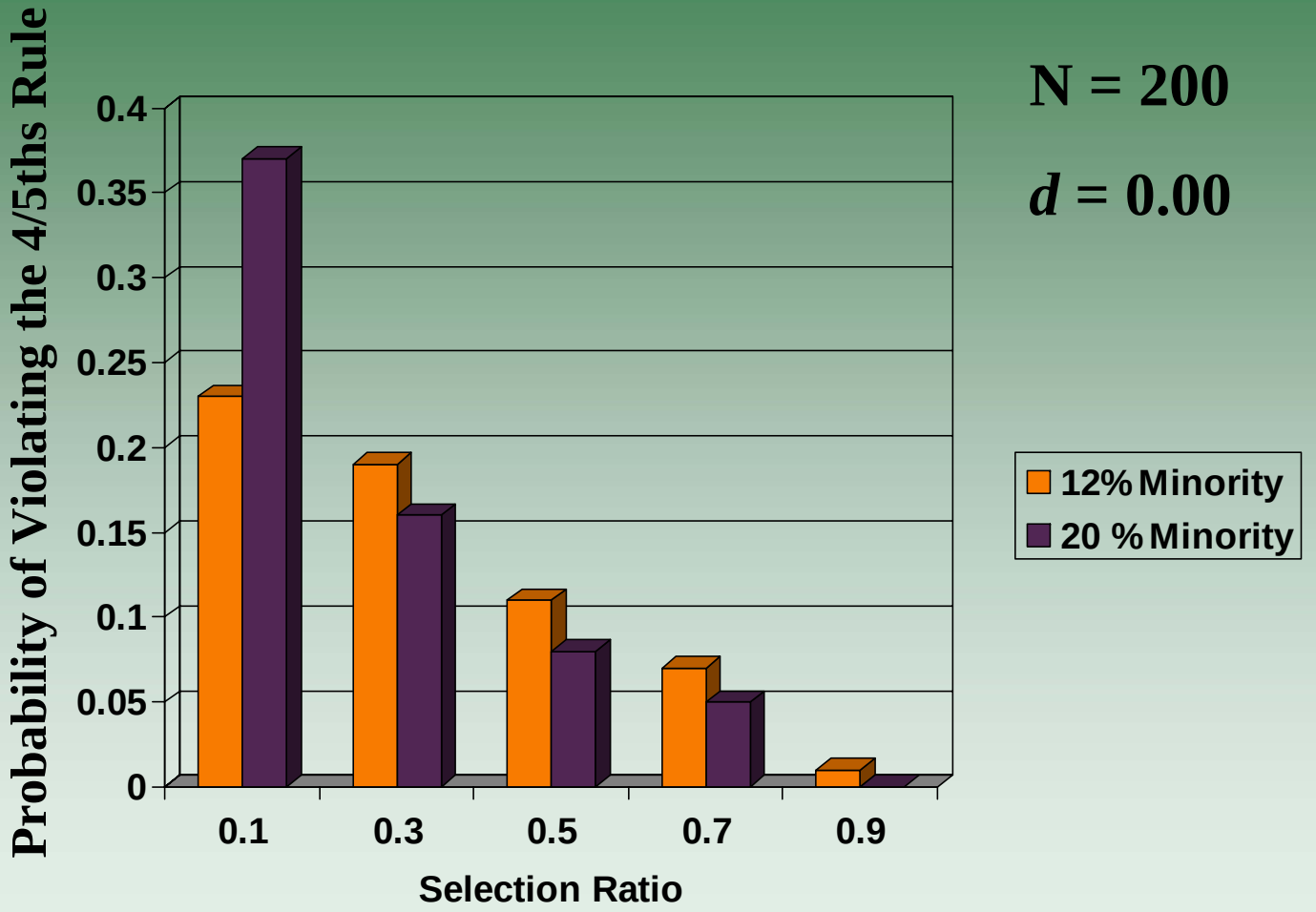
#4: Context matters!!!

- **Don’t forget complexity.**
  - **Reading level**
  - **Math level**
  - **Skills/Abilities level**
- **Context is also important in evaluating adverse impact.**
  - **Adverse impact is influenced by factors unrelated to the test.**
  - **Consider the sample—Applicant samples are better indicators of adverse impact than incumbent samples.**

Example—Adverse impact is influenced by factors unrelated to the test

- **Total Sample Size**
- **Number of Minorities in the Sample**
- **Selection Ratio**
- **Correlation Between Predictors**

Example: AI Ratios From a Single-Hurdle Selection System



N = 200

***d* = 0.00**

Roth, Bobko, & Switzer, 2006

Example—Consider the sample when evaluating adverse impact.

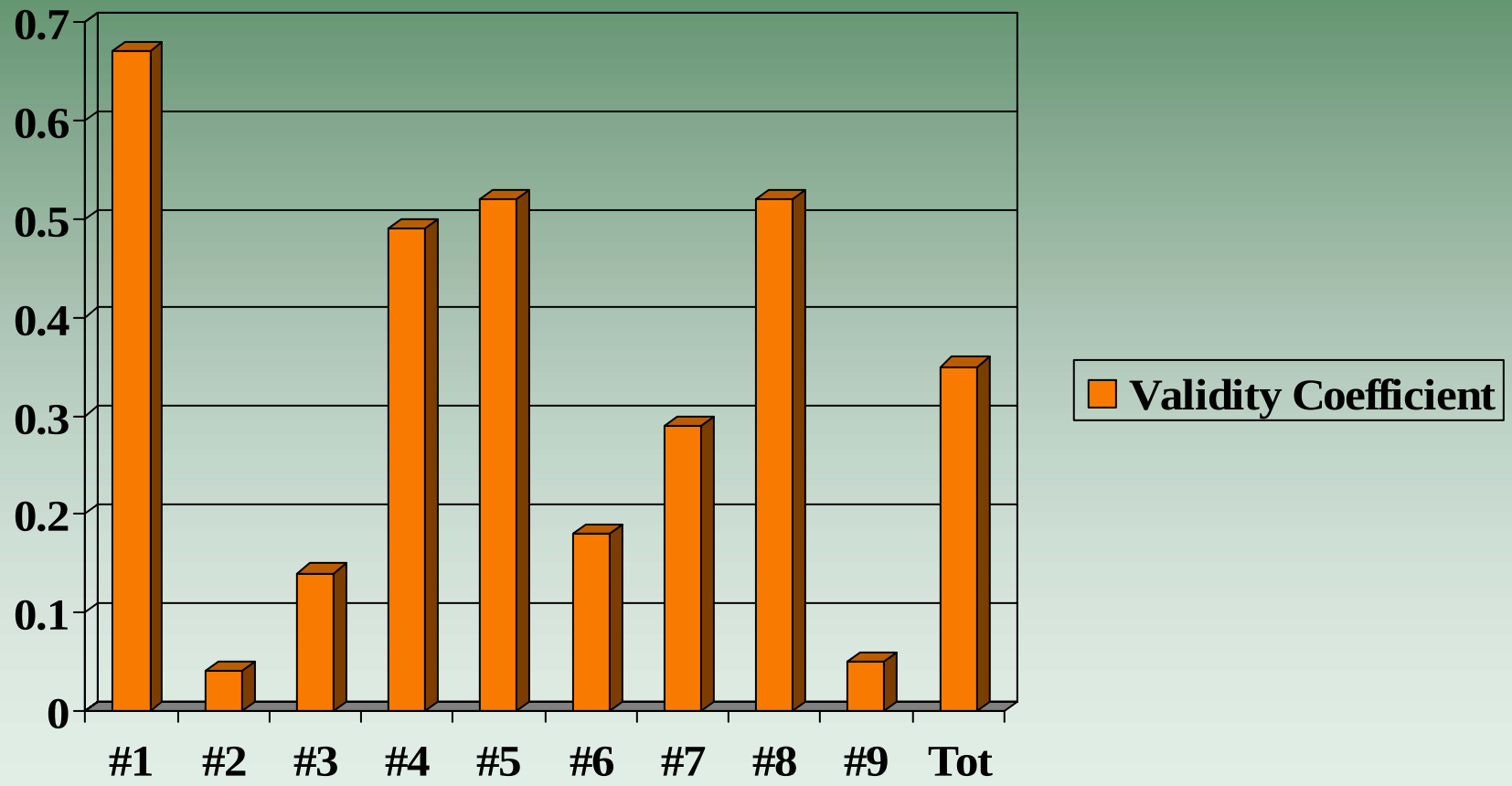
- **Applicant samples generally demonstrate higher adverse impact than incumbent samples.**

	White-Black SD-Difference in Validation Sample	White-Black SD-Difference in Applicant Sample
Test #1	-.10	.50
Test #2	.15	.54
Test #3	-.66	.24
Test #4	-.16	.41
Test #5	-.13	.69

#5: Beware of small samples

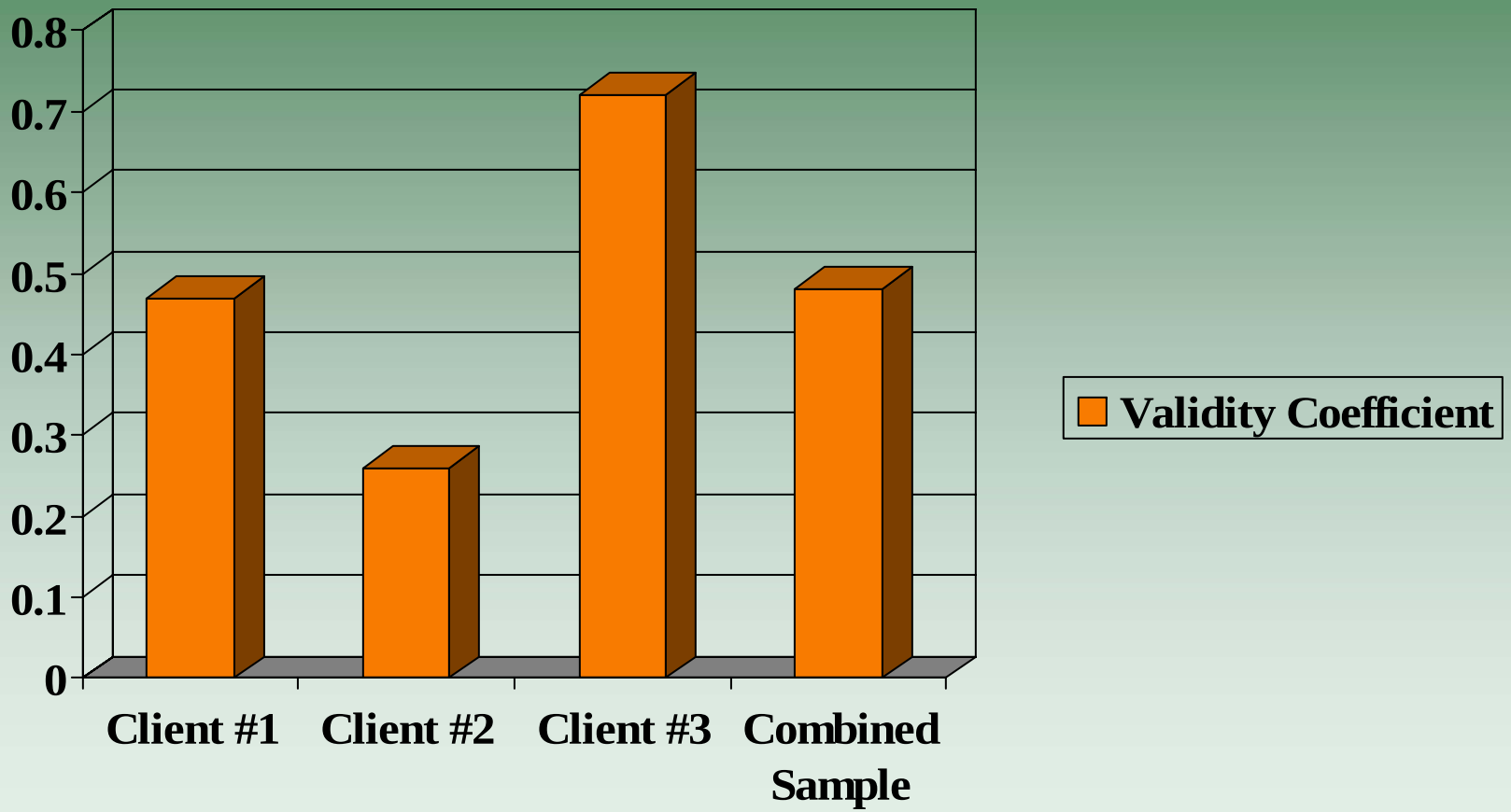
- **“Ignoring sampling error leads to disastrous results in the area of personnel selection.” (Hunter & Hunter, 1984)**
- **Sampling error occurs due to only sampling part of the entire population**
  - **Single studies and/or small samples are not definitive.**
  - **Results from single studies and/or small samples are not robust.**

Example: Sampling Error—Smaller Samples



Single Test Validated in Multiple Samples (All samples > 20 participants)

Example: Sampling Error- Larger Samples



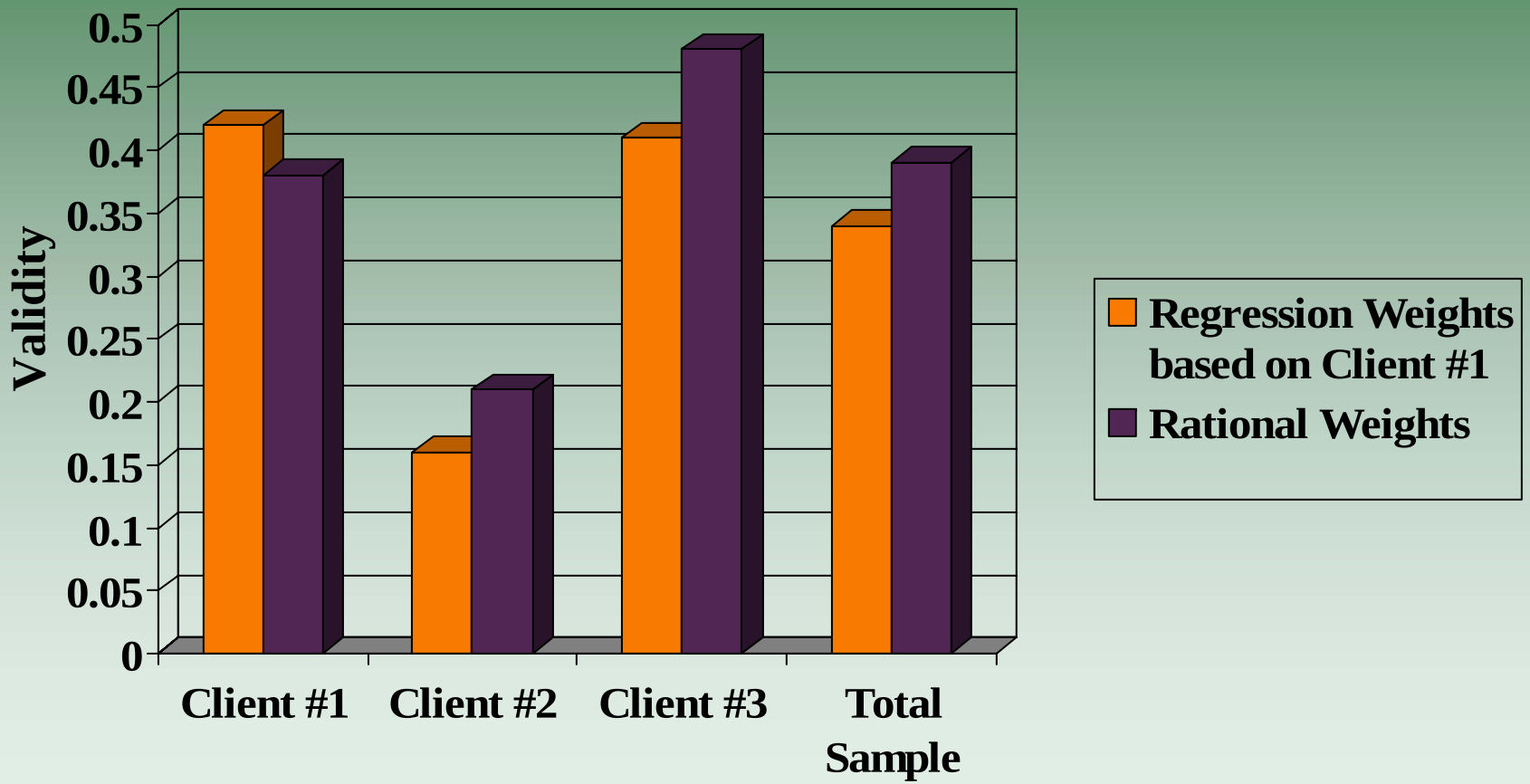
Next Generation Firefighter/EMS Written Aptitude Test

(All Samples > 65 participants)

#5: Beware of small samples

- **Capitalizing on chance can result in misleading validity coefficients.**
- **Capitalization on chance can occur when:**
  - **Final items on test are determined based on validation sample.**
  - **Test weights are determined based on validation sample.**
- **You should expect lower validity coefficients in the future under these circumstances.**

Example: Capitalization on Chance

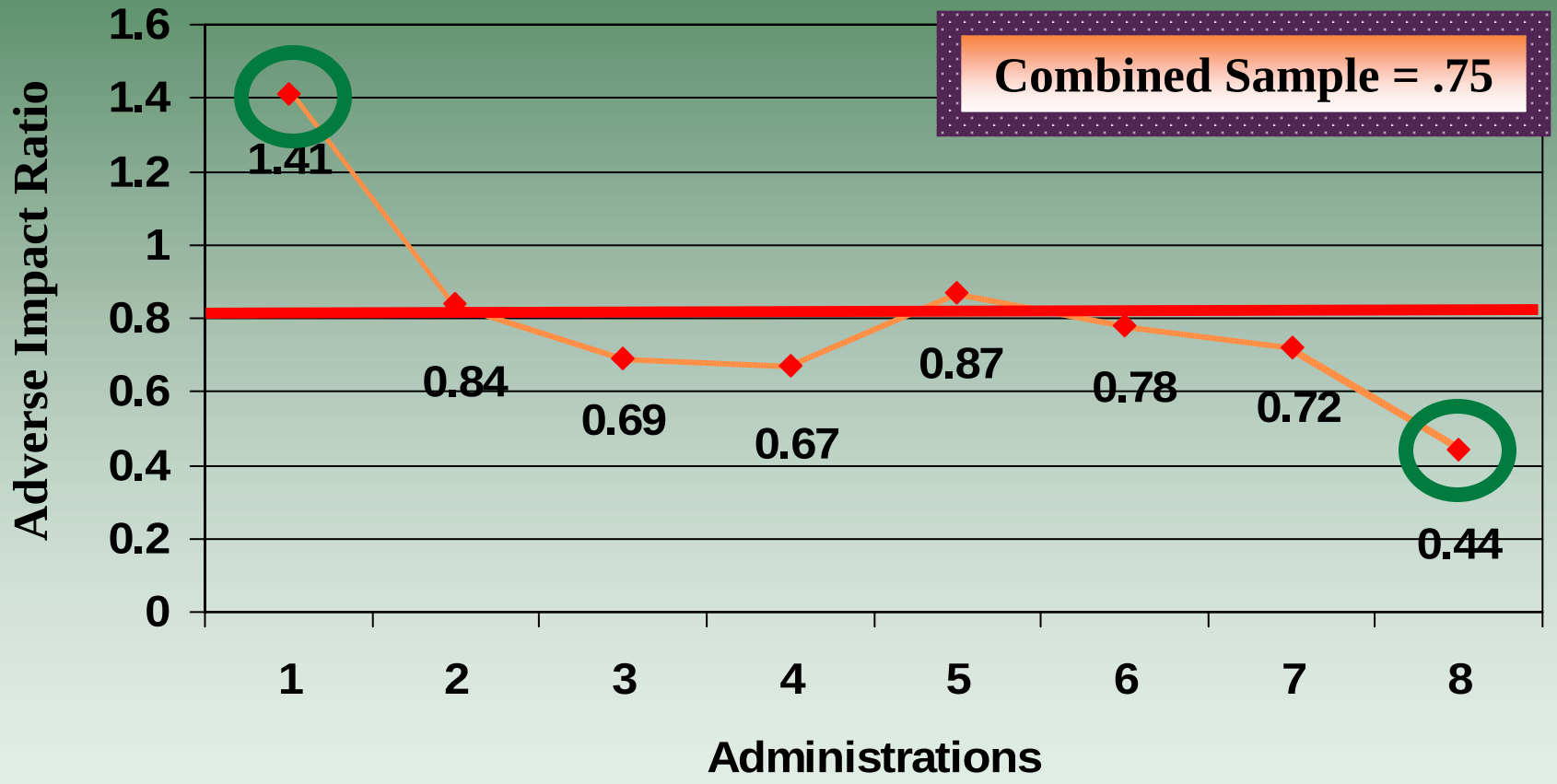


#5: Beware of small samples

- **Single studies/small samples can also result in misleading adverse impact ratios.**
  - **The 4/5ths rule is not AI. It is an indicator of underlying AI—"The 4/5ths rule merely establishes a numerical basis for drawing an initial inference and for requiring additional information" (*Uniform Guidelines, Questions & Answers*)**

- **AI ratios can vary substantially over different administrations.—Again, results from single studies and/or small samples are not definitive or robust.**

Example: One Client's AI Ratios Over Multiple Administrations



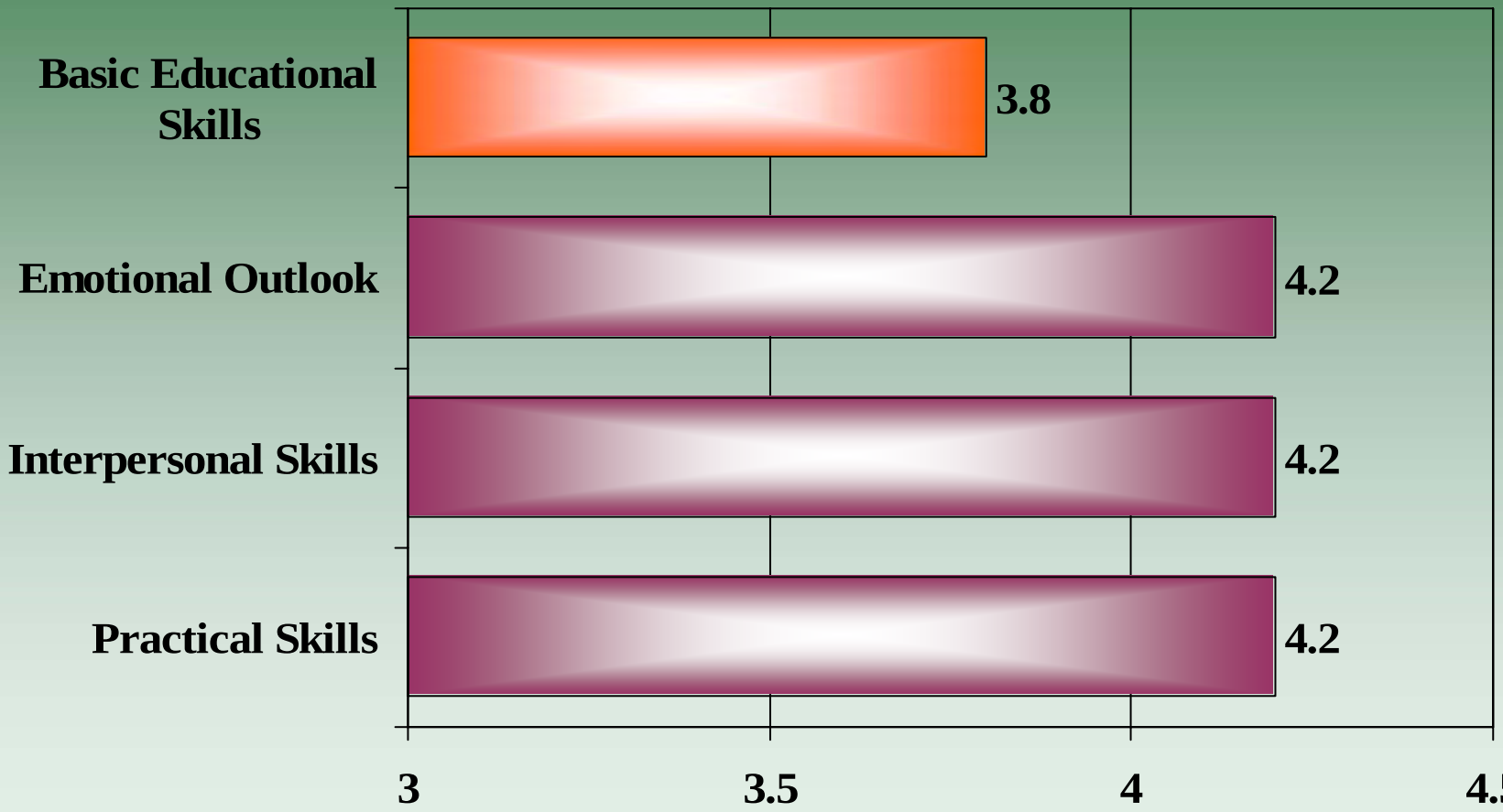
#5: Beware of small samples

- **When evaluating samples:**
  - **More weight should be given to evidence from multiple samples—Cross validation.**
  - **More weight should be given to larger samples.**
  - **More weight should be given to representative samples.**
  - **More weight should be given to results from studies that are developed and weighted using rational models.**

#6: Don’t forget the O’s

- **The concept of KSAs has been expanded to KSAOs**
  - **O’s = Other Characteristics**
    - **Judgment & Common Sense**
    - **Interpersonal Skills**
    - **Emotional Skills**
    - **Leadership**
    - **Personality traits or temperaments**
    - **Interests**
- **Defining a broader job domain can result in higher validity and lower adverse impact.**

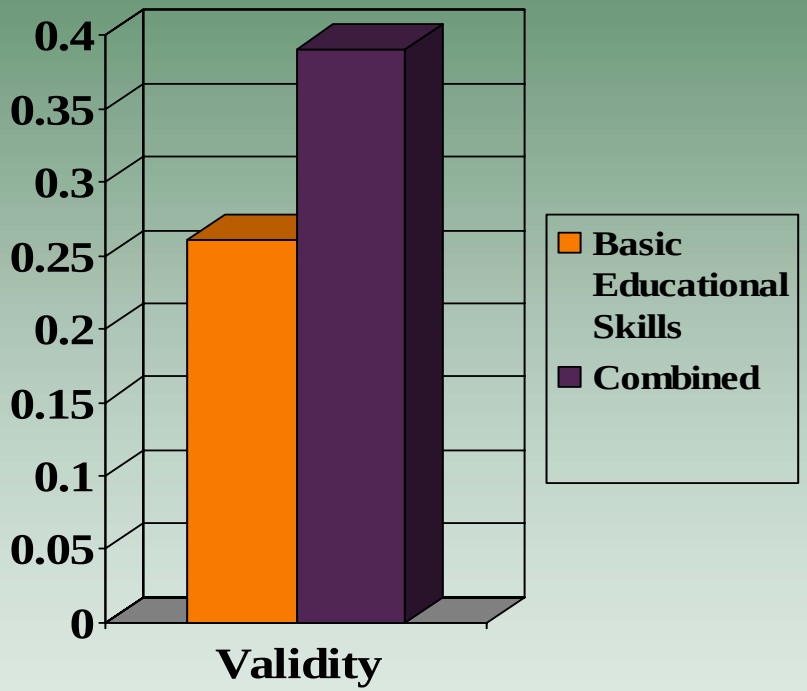
Example: KSAO Importance Ratings for Firefighter Position

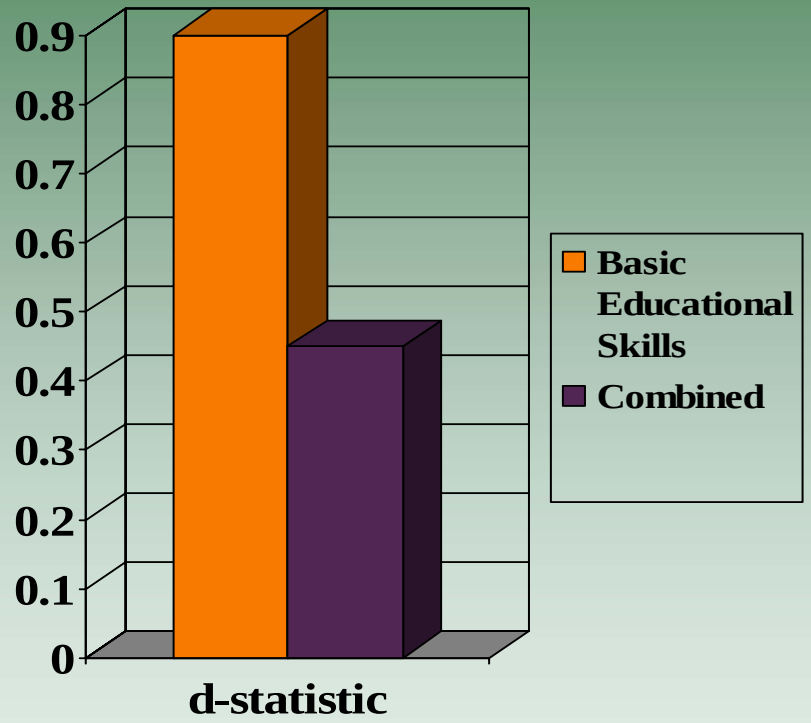


Very Important
Essential for successful performance of the job

Critically Important
Failure to perform results in extreme negative consequences

Example: Broad Assessments Can Increase Validity & Reduce Adverse Impact





Combined includes Interpersonal Skills, Emotional Outlook, & Practical Skills

#6: Don’t forget the O’s

- **Using broader assessments early in the process can result in substantially better hires.**
  - Some agencies administer a narrow test (e.g., basic educational skills) in the first stage and measure a broader range of skills in a later stage (e.g., interview).
  - This strategy will screen out individuals who are more complete candidates and would be superior employees.
  - Measuring a broad range of skills can increase the validity (i.e., the quality of the candidate pool) and minimize the AI of your first stage (as well as your total process).
  - Measuring a broad range of skills early in your process can also reduce the cost of later steps.

Example: Which Candidate Would be the Best Hire?

	Basic Educational Skills	Interpersonal	Emotional Outlook	Practical
Candidate A	87	60	60	60
Candidate B	85	70	70	70
Candidate C	83	90	90	90

Example: Advantage of Measuring a Broad Range of Skills Early in Process

Selection Ratio	AI Ratio-Cognitive Screen	AI Ratio-Complete Model	% of Top Candidates Screened Out by Cognitive Screen
.20	.32	.32	68%
.40	.37	.49	35%
.60	.52	.65	23%
.80	.63	.85	12%

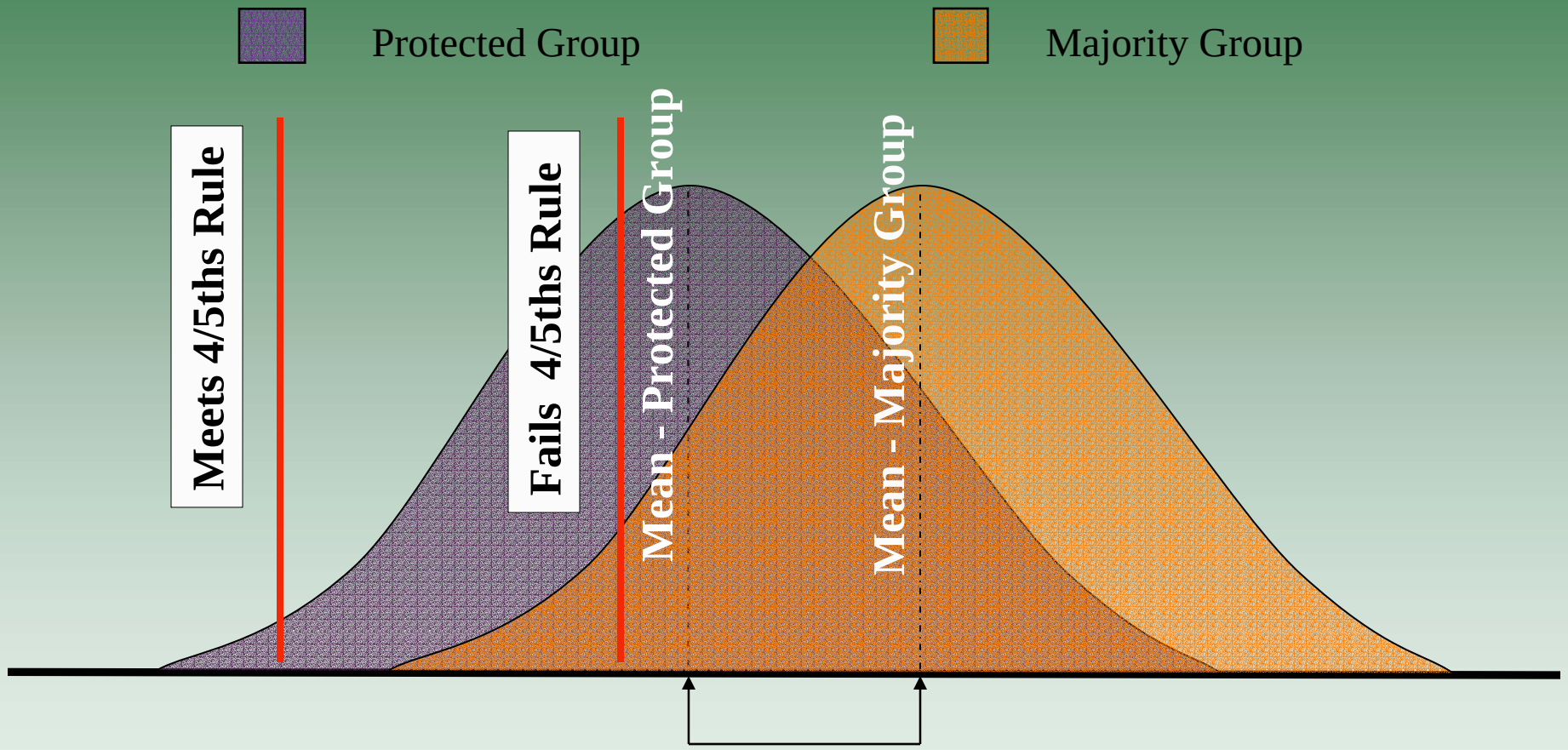
#7: Remember the Evaluate the Pass Point

- Adverse impact ratios are dependent upon pass points.
  - Adverse Impact Ratio—A substantially different ***rate of selection*** is indicated when the ***selection rate*** for a protected group is less than 4/5ths (80%) of the ***selection rate*** for the group with the highest selection rate.

- Changing the pass point results changes the AI Ratio.

- Make sure the pass point used by test provider when evaluating adverse impact is similar to your expected pass point.

Example: Adverse impact ratios are dependent on pass points

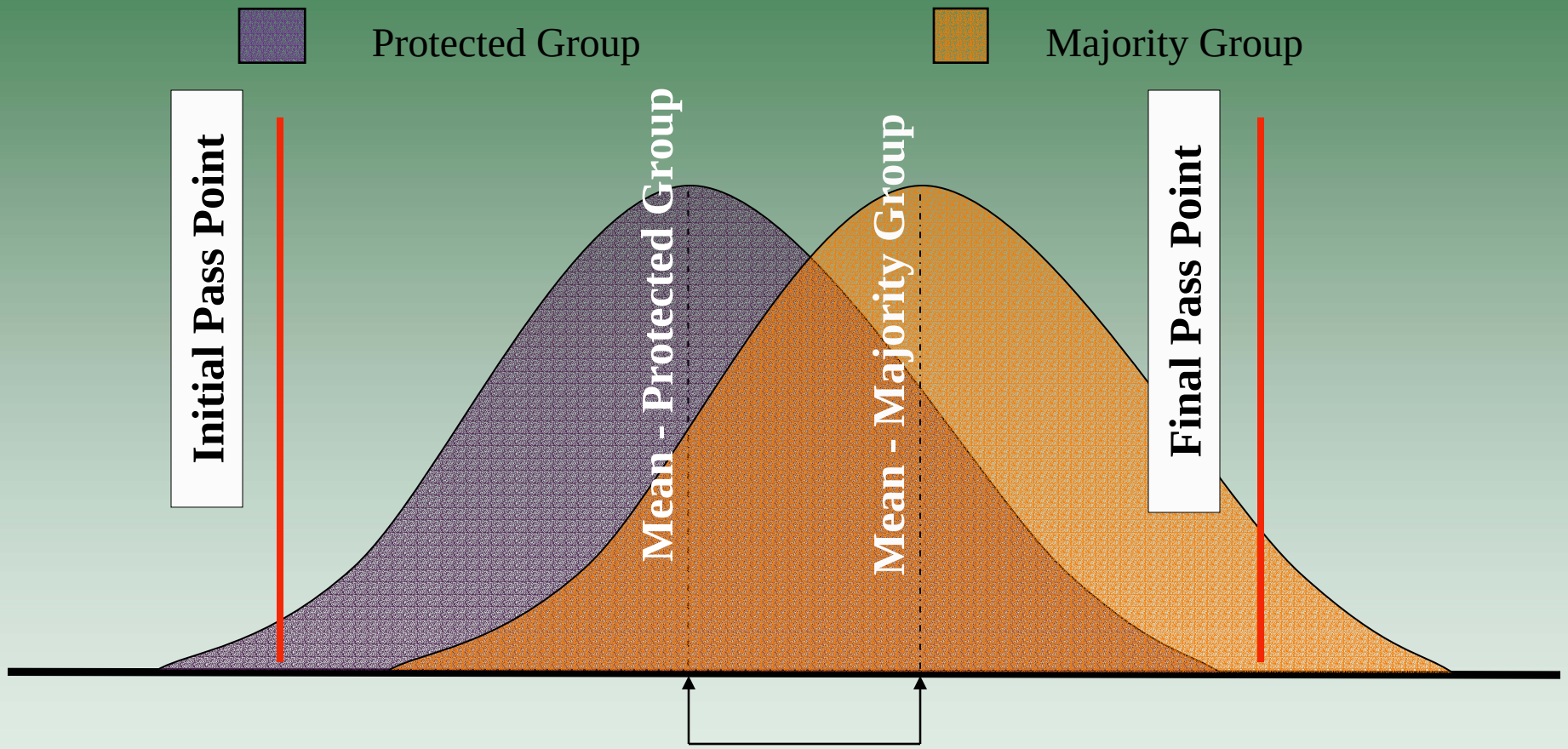


#7: Remember to evaluate the pass point

- **Remember that your process may have multiple pass points.**
  - **Those that pass test**
  - **Those that are ultimately hired**

- **Although your initial pass point may meet the 4/5ths rule, the rank-order is a critical consideration.**

Example: Rank order is critical consideration



Example: Rank order impacts AI Ratio of your ultimate pass point

Rank	Score	Race
1	92	W
2	88	W
3	87	W
4	86	B
5	81	W
6	80	W
7	79	W
8	78	B
9	77	W
10	76	B
11	75	B
12	72	W
13	71	W
14	70	W

Selection Process Results

10 of 70 W pass
4 of 30 B pass

W pass ratio = 14.3 %
B pass ratio = 13.3 %

AI Ratio = 0.93

Hires: 4 W, 1 B
Hire ratios: W = 5.7%, B = 3.3%

Hire ratio AI = 0.58

Conclusion: Adverse Impact

Hires: 7 W, 3 B
Hire ratios: W = 10%, B = 10%

Hire ratio AI = 1.0

Conclusion: No Adverse Impact

CWH

7 Critical Considerations When Evaluating Selection Tests

1. **Don’t let the tail wag the dog—Take control of your process.**
2. **There is no such thing as a valid test.**
3. **Not all validity evidence is equal.**
4. **Context matters!!!**
5. **Beware of small samples.**
6. **Don’t forget the O’s**
7. **Remember to evaluate the pass point.**