

---

# Written Test Items: Toss 'Em or Keep 'Em?:

What you need to know about removing versus keeping  
“bad” test items.

James H. Killian, Ph.D.

Michael D. Blair, M.S.

***CWH Management Solutions***

---

# Introduction

---

- Rationale for investigation
  - Group differences in employee testing
    - Adverse Impact
    - *SD* Difference
  - Case Studies
  - Conclusions & Recommendations
-

# Rationale for Investigation

---

- Personnel testing is commonplace in most public entities (e.g., fire, police, federal)
  - Concern for “good tests”
    - Potential for bias, poor test quality & associated legal repercussions
-

# The Good, Bad & Ugly Item

---

- Two Main Routes by Which “bad” Items May be Modified After Test is Completed
    - Appealed Items
    - Items with Poor Psychometric Qualities
  - Two Schools of Thought
    - Drop Item and Re-score Test
    - Give everyone credit for the item
  - No Research to Support Either Argument!
-

# Adverse Impact (AI)

---

- One group (e.g., White) selected at a higher rate than another other group (e.g., Black)
  - Courts use 4/5 rule (80% rule)
    - The selection ratio for a minority group should not be less than 80% of the rate for the majority group
-

# Adverse Impact

---

- The 4/5 “rule” is not absolute – it is a guideline
  - *SD* Difference, or *d*-statistic, can be conceptualized as the standardized difference between subgroups
-

# SD Difference formula

---

$$d = \sqrt{V_W}$$
$$V_W = \frac{[(N_E - 1) * V_E] + [(N_C - 1) * V_C]}{(N_E - 1) + (N_C - 1)}$$

Where:

$N_E$  = Majority Sample Size

$V_E$  = Majority Group Variance

$N_C$  = Minority Sample Size

$V_C$  = Minority Group Variance

$V_W$  = Pooled Within Group Variance

The difference in group (i.e., Black, White) means, divided by the pooled standard deviation of each group (i.e., sample-weighted).

---

# What you need to know about *SD* Difference

---

- An *SD* Diff statistic of 0 indicates that the Majority and Minority groups scored the same
  - A negative *SD* Diff statistic indicates that, on the whole, the Minority group scored better on the whole than the Majority\*
  - Typical written tests result in *SD* Diff statistics that range from 1.00 to 1.40, indicating that Whites scored better than Minorities on the whole
-



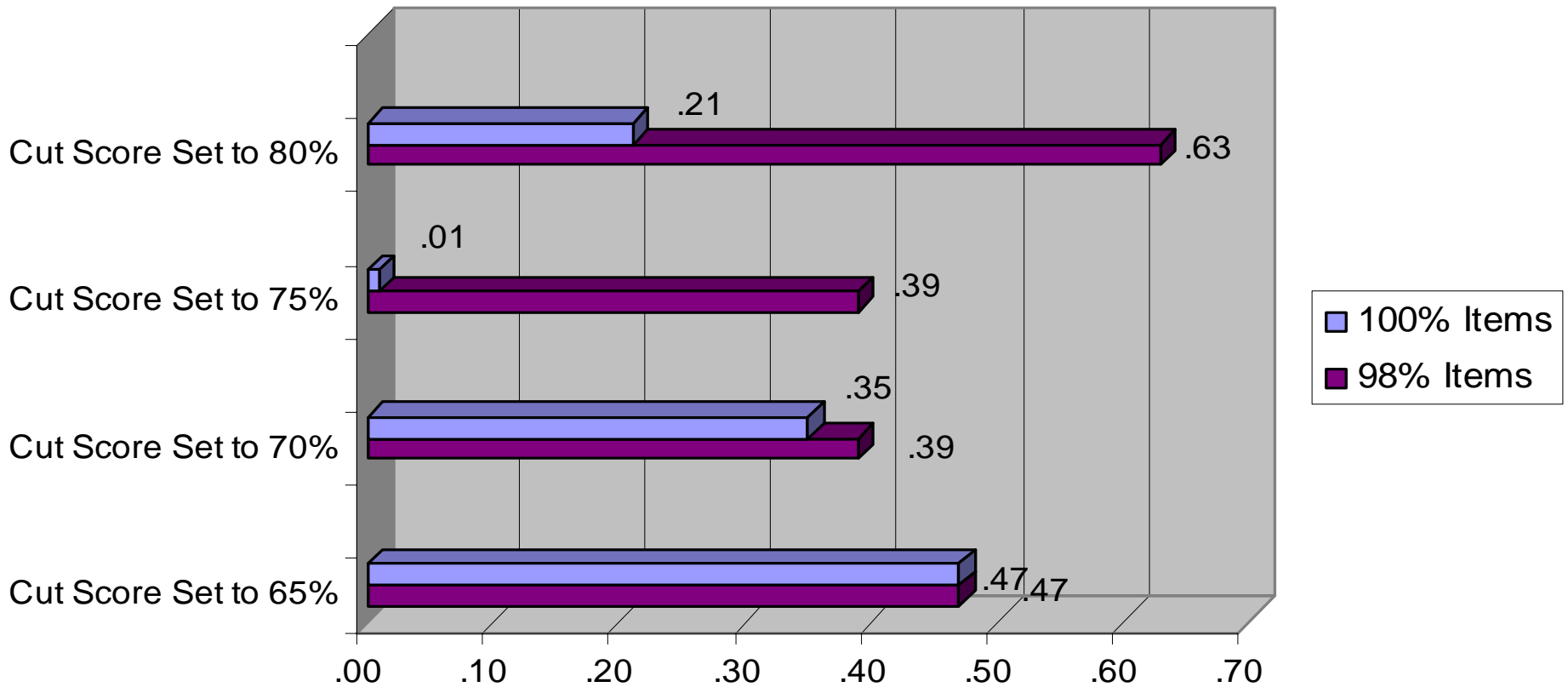
# General Hypothesis

---

- Removing Items and Re-Scoring Tests Leads to Increased Group Differences Between Whites and Minorities
-

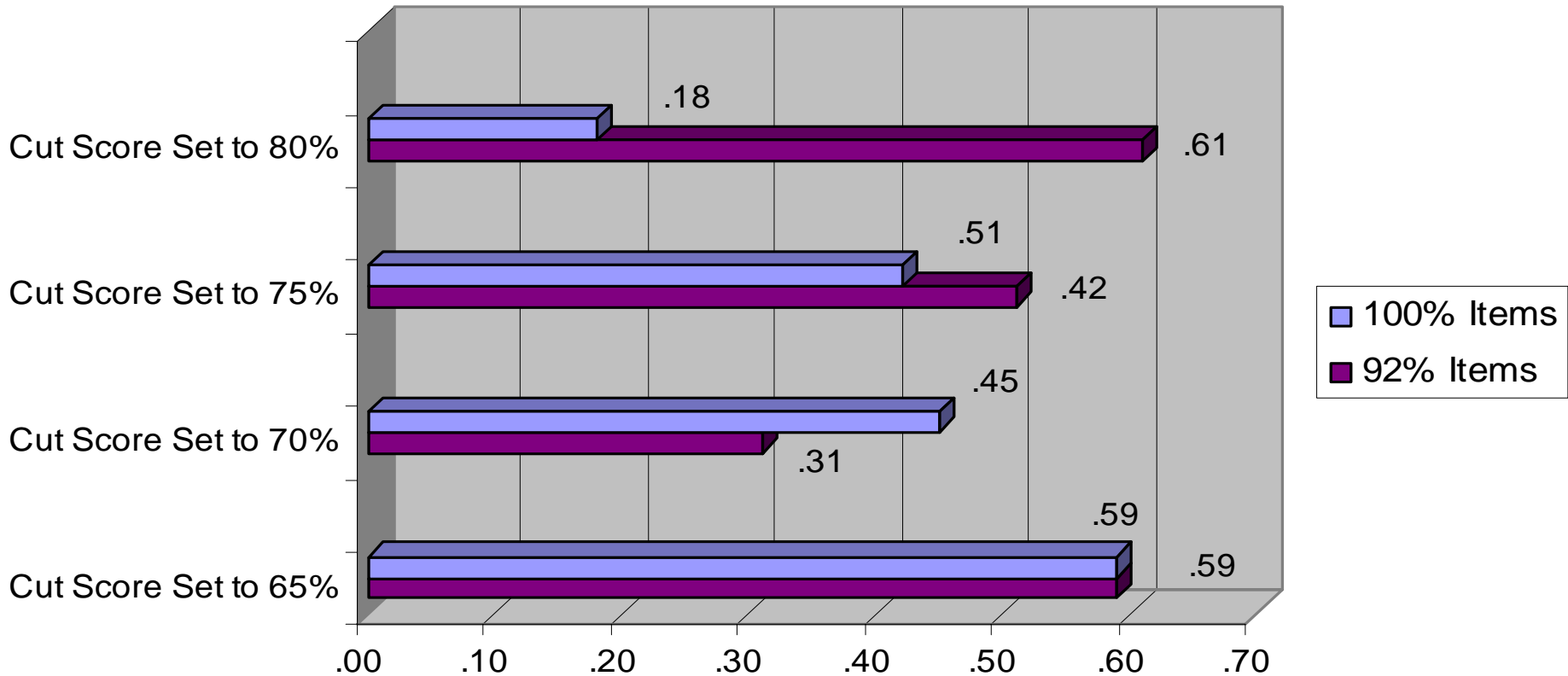
# Sample Case #1

## SD Differences Presented at Cut Scores



# Sample Case #2

## SD Differences Presented at Cut Scores



# Results

---

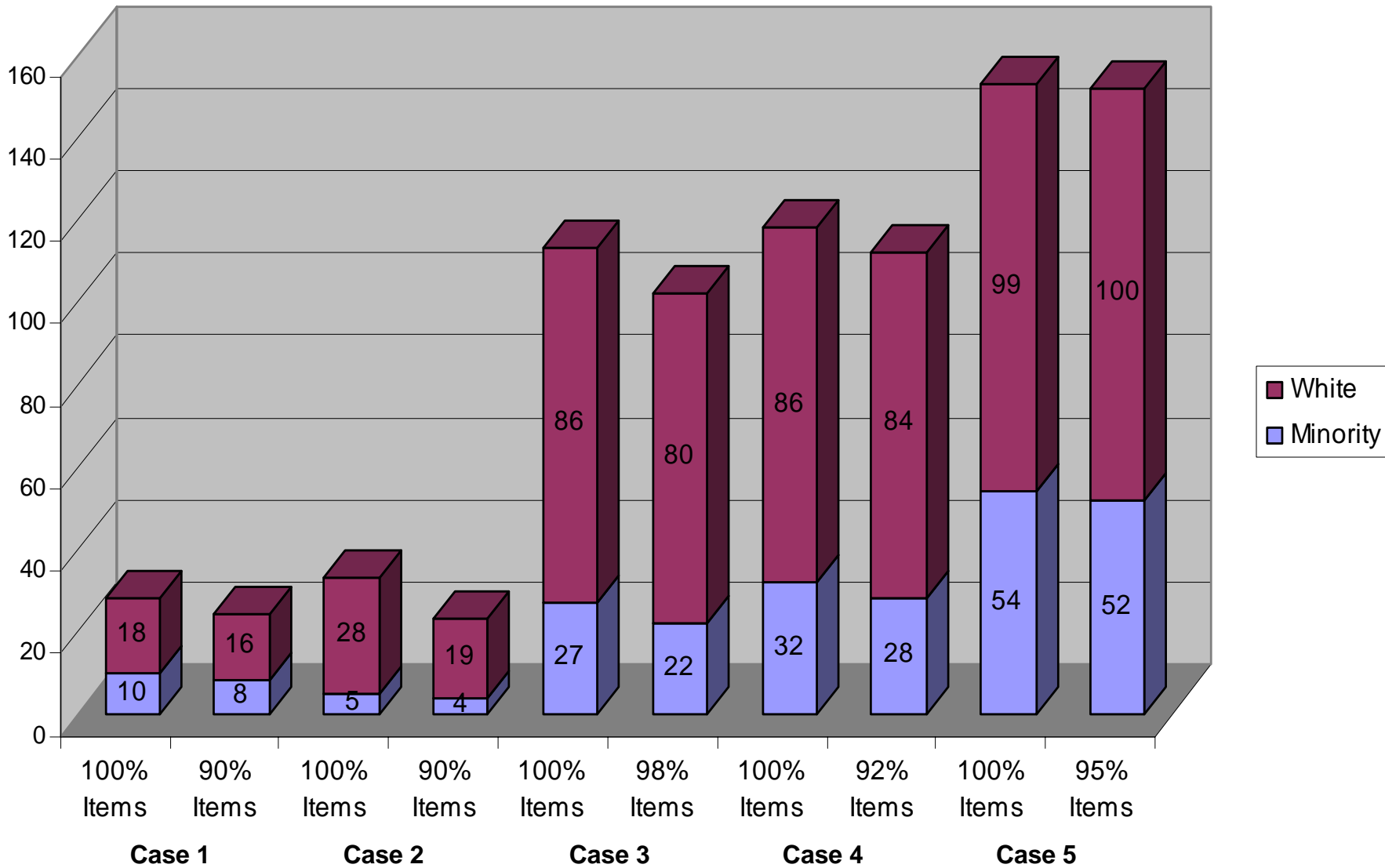
- Psychometric approach does not support one method over the other
    - Results were mixed
  - Practical results were somewhat different and more useful
-

# A Practical Approach

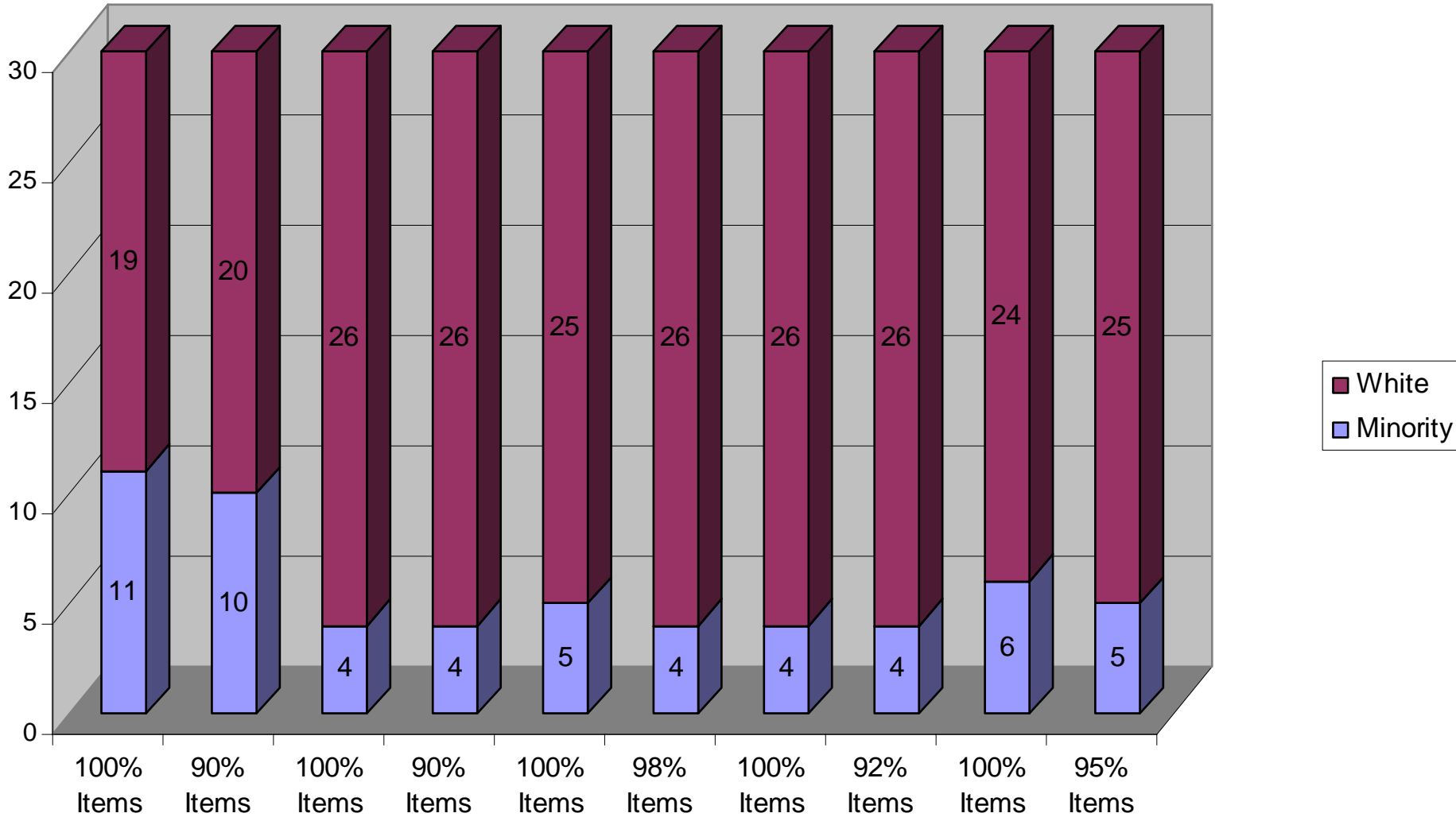
---

- Simple review of samples based on two common criteria
    - 70% Written Test Cut Score
    - Top 30 candidates
-

# Candidates Scoring 70% or Greater Across Five Written Tests



# Top 30 Candidates Written Test Candidates



**Case 1**

**Case 2**

**Case 3**

**Case 4**

**Case 5**

# Conclusions & Suggestions

---

- Psychometric approach does not support one method (dropping/keeping) over the other.
  - Practical approach appears to support giving candidates item credit over removing “bad” items and re-scoring.
    - Results in greater number of minorities at the upper end of Written Test promotional lists
-



# Limitations

---

- Current confound due to internal test data
    - SD differences ranged from .00 to .77, compared to 1.00 to 1.40 typically found
    - Results may not be completely applicable to typical technical job knowledge or mechanical reasoning tests
    - Limited generalizability
-

# Future Research

---

- Acquire strictly technical/mechanical reasoning data to analyze and compare
  - Research entry-level tests
  - Conduct Monte Carlo Simulation
  - Track Majority/Minority Item Appeals
-

---

# Questions & Comments

---

---

# Additional Materials

---

# Adverse Impact Example 1

---

| Number of Applicants | Number Hired | Selection Ratio | Adverse Impact             |
|----------------------|--------------|-----------------|----------------------------|
| 100<br>WHITE         | 20           | .20             | $.10 / .20 =$<br>.50 (50%) |
| 50 BLACK             | 5            | .10             |                            |

---

# Adverse Impact Example 2

---

| Number of Applicants | Number Hired | Selection Ratio | Adverse Impact             |
|----------------------|--------------|-----------------|----------------------------|
| 100<br>WHITE         | 20           | .20             | $.16 / .20 =$<br>.80 (80%) |
| 50 BLACK             | 8            | .16             |                            |

---

# Case Study 1

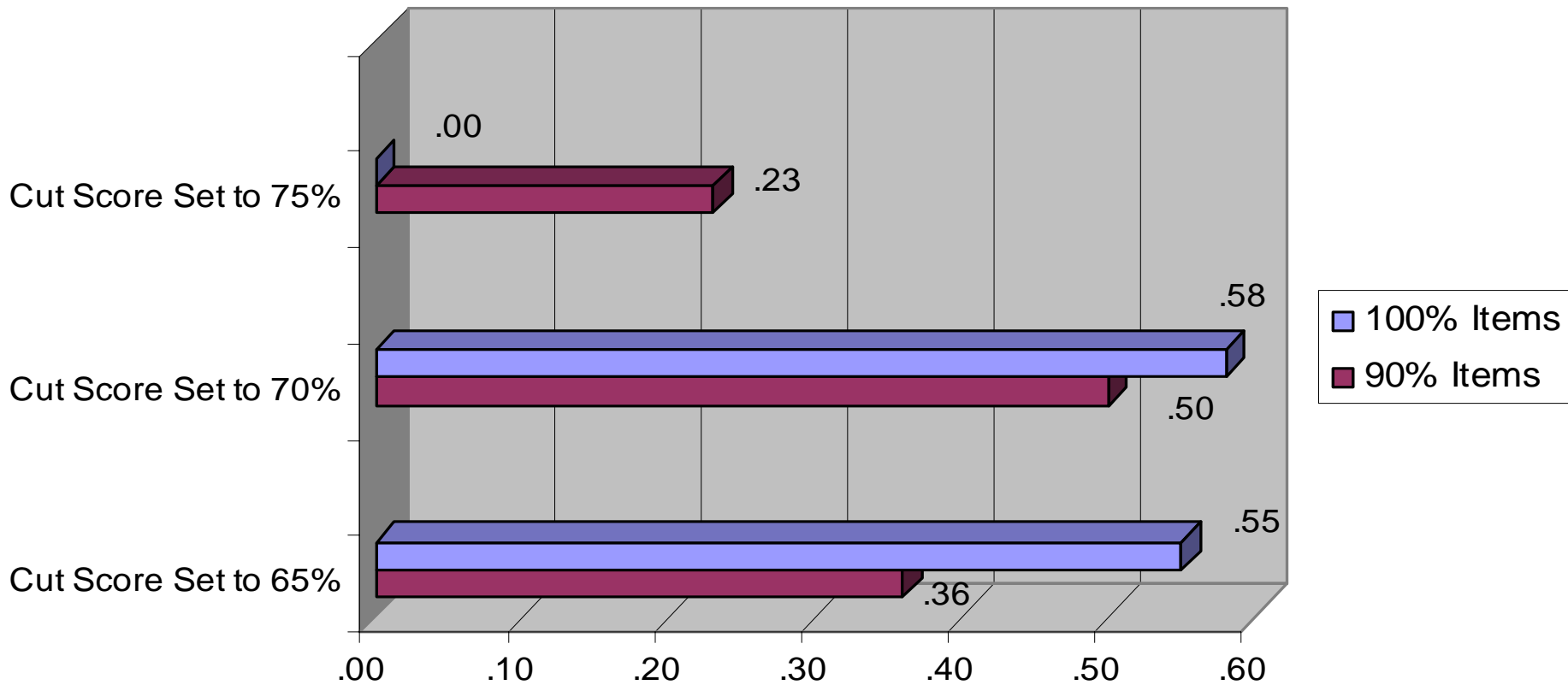
|                      |          | 100% Items |          |           |           |          | After Appealed Items Are Removed* |          |           |           |          |
|----------------------|----------|------------|----------|-----------|-----------|----------|-----------------------------------|----------|-----------|-----------|----------|
|                      |          | <i>N</i>   | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> | <i>N</i>                          | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> |
| No Cut Score         | White    | 21         | 74.38    | 5.14      | --        | --       | 21                                | 72.65    | 5.56      | --        | --       |
|                      | Minority | 13         | 71.92    | 4.37      | --        | --       | 13                                | 69.91    | 5.18      | --        | --       |
| Cut Score Set to 80% | White    |            |          |           | #####     | --       |                                   |          |           | ####      | --       |
|                      | Minority |            |          |           |           |          |                                   |          |           |           |          |
| Cut Score Set to 75% | White    | 10         | 78.50    | 3.06      | .32       | -.17     | 7                                 | 78.89    | 2.64      | .46       | .23      |
|                      | Minority | 2          | 79.00    | 0.00      |           |          | 2                                 | 78.33    | 0.79      |           |          |
| Cut Score Set to 70% | White    | 18         | 75.83    | 3.85      | .90       | .58      | 16                                | 74.86    | 4.10      | .81       | .50      |
|                      | Minority | 10         | 73.80    | 2.82      |           |          | 8                                 | 72.92    | 3.46      |           |          |
| Cut Score Set to 65% | White    | 20         | 74.95    | 4.55      | .97       | .55      | 19                                | 73.68    | 4.70      | .85       | .36      |
|                      | Minority | 12         | 72.58    | 3.82      |           |          | 10                                | 72.11    | 3.49      |           |          |

**Note:** Ten (10) items were removed prior to rescoring and converting each candidate's score to a percentile.

N = 34

# Case Study 1

## SD Differences Presented at Cut Scores





# Case Study 2

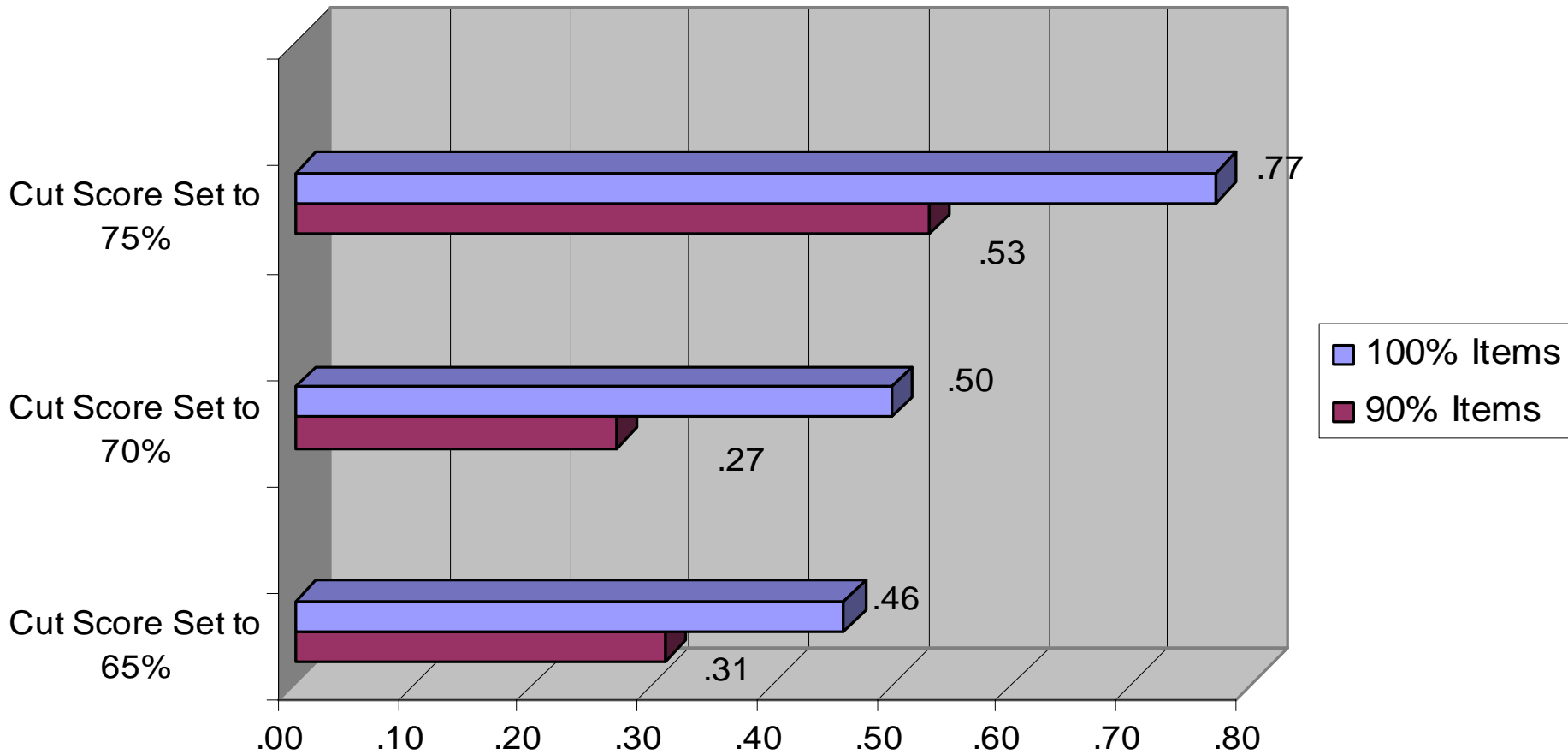
|                      |          | 100% Items |          |           |           |          | After Appealed Items Are Removed* |          |           |           |          |
|----------------------|----------|------------|----------|-----------|-----------|----------|-----------------------------------|----------|-----------|-----------|----------|
|                      |          | <i>N</i>   | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> | <i>N</i>                          | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> |
| No Cut Score         | White    | 36         | 76.77    | 6.83      | --        | --       | 36                                | 70.96    | 6.45      | --        | --       |
|                      | Minority | 12         | 69.55    | 7.55      | --        | --       | 12                                | 65.07    | 7.29      | --        | --       |
| Cut Score Set to 80% | White    | 12         | 84.27    | 2.84      | .25       | --       | 12                                | 77.95    | 2.95      | .25       | --       |
|                      | Minority | 1          | 80.37    | .         |           |          | 1                                 | 74.75    | .         |           |          |
| Cut Score Set to 75% | White    | 22         | 81.22    | 4.06      | .55       | .77      | 22                                | 75.07    | 4.15      | .55       | .53      |
|                      | Minority | 4          | 78.27    | 1.59      |           |          | 4                                 | 72.98    | 1.52      |           |          |
| Cut Score Set to 70% | White    | 28         | 79.54    | 4.91      | .54       | .50      | 19                                | 73.41    | 4.96      | .63       | .27      |
|                      | Minority | 5          | 77.20    | 2.77      |           |          | 4                                 | 72.12    | 2.33      |           |          |
| Cut Score Set to 65% | White    | 44         | 76.77    | 6.83      | .55       | .46      | 34                                | 70.96    | 6.45      | .44       | .31      |
|                      | Minority | 8          | 73.71    | 5.25      |           |          | 5                                 | 69.07    | 4.70      |           |          |

**Note:** Ten (10) items were removed prior to rescoring and converting each candidate's score to a percentile.

N = 48

# Case Study 2

## SD Differences Presented at Cut Scores



# Case Study 3

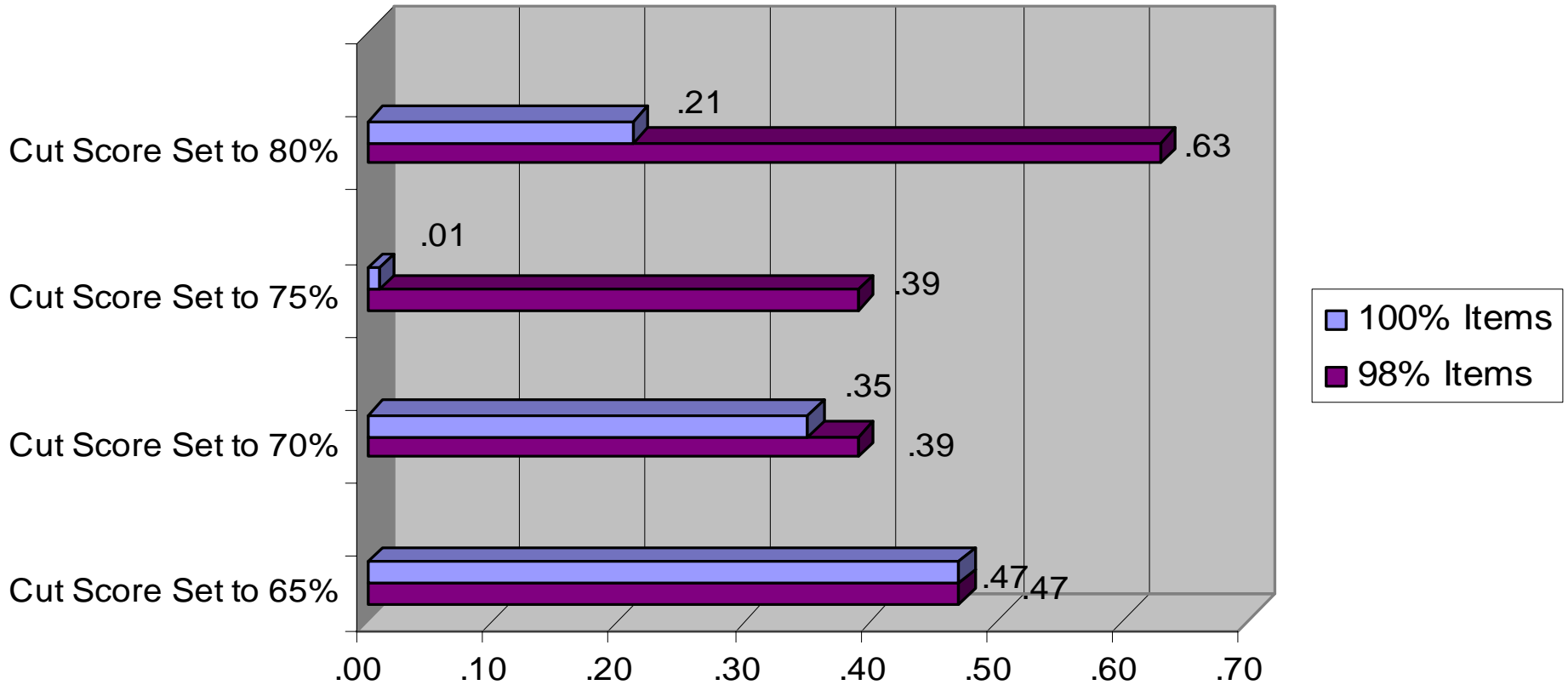
|                      |          | 100% Items |          |           |           |          | After Appealed Items Are Removed* |          |           |           |          |
|----------------------|----------|------------|----------|-----------|-----------|----------|-----------------------------------|----------|-----------|-----------|----------|
|                      |          | <i>N</i>   | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> | <i>N</i>                          | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> |
| No Cut Score         | White    | 121        | 73.36    | 6.97      | --        | --       | 121                               | 72.77    | 7.14      | --        | --       |
|                      | Minority | 59         | 69.10    | 7.13      | --        | --       | 59                                | 67.99    | 6.99      | --        | --       |
| Cut Score Set to 80% | White    | 20         | 83.15    | 3.28      | .41       | .21      | 13                                | 84.69    | 3.25      | .32       | .63      |
|                      | Minority | 4          | 82.50    | 2.38      |           |          | 2                                 | 82.65    | 2.89      |           |          |
| Cut Score Set to 75% | White    | 57         | 79.19    | 3.72      | .43       | .01      | 48                                | 79.61    | 3.73      | .47       | .39      |
|                      | Minority | 12         | 79.17    | 2.98      |           |          | 11                                | 78.20    | 2.75      |           |          |
| Cut Score Set to 70% | White    | 86         | 76.77    | 4.63      | .64       | .35      | 80                                | 76.76    | 4.64      | .56       | .39      |
|                      | Minority | 27         | 75.19    | 4.24      |           |          | 22                                | 75.00    | 3.90      |           |          |
| Cut Score Set to 65% | White    | 110        | 74.66    | 5.76      | .84       | .47      | 107                               | 74.37    | 5.81      | .77       | .47      |
|                      | Minority | 45         | 72.02    | 5.22      |           |          | 40                                | 71.76    | 4.70      |           |          |

**Note:** Two (2) items were removed prior to rescoring and converting each candidate's score to a percentile.

N = 180

# Case Study 3

## SD Differences Presented at Cut Scores



# Case Study 4

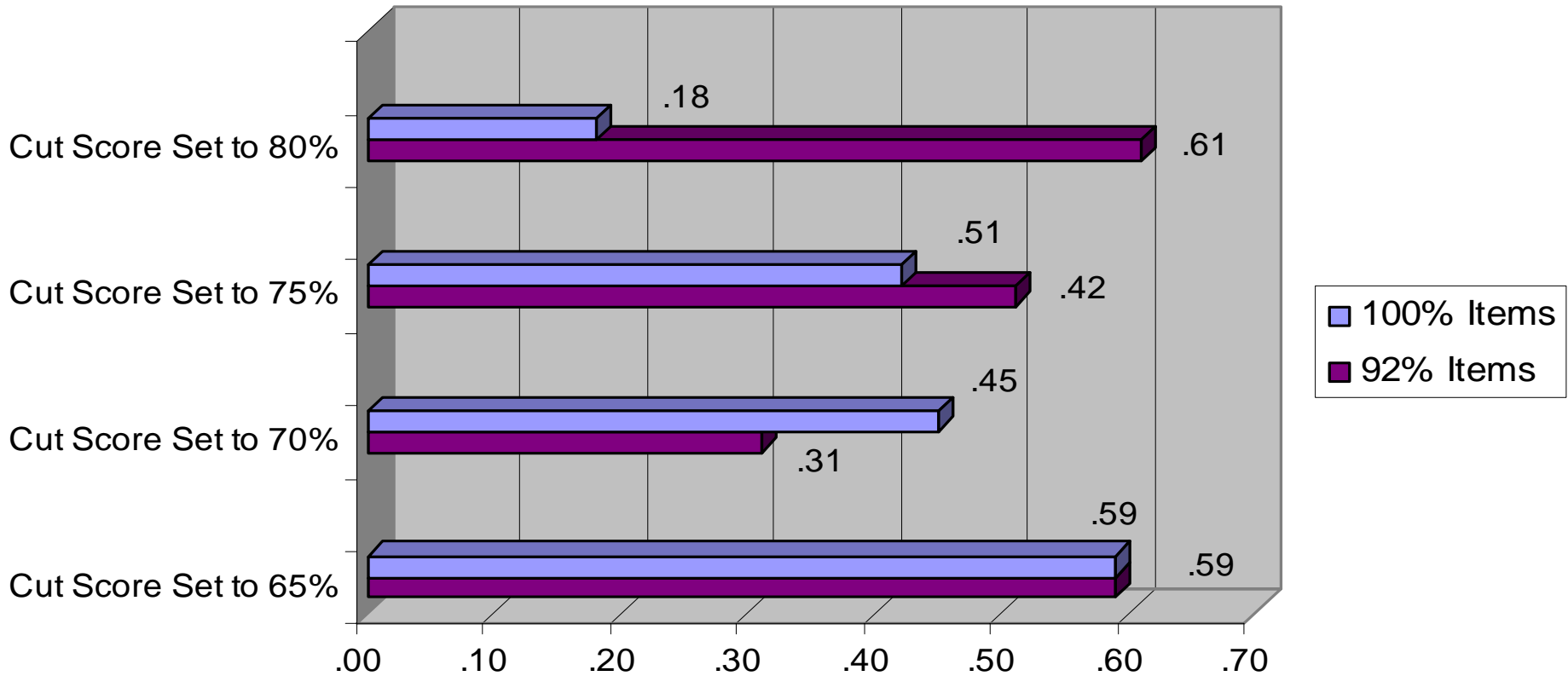
|                      |          | 100% Items |          |           |           |          | After Appealed Items Are Removed* |          |           |           |          |
|----------------------|----------|------------|----------|-----------|-----------|----------|-----------------------------------|----------|-----------|-----------|----------|
|                      |          | <i>N</i>   | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> | <i>N</i>                          | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> |
| No Cut Score         | White    | 104        | 79.82    | 9.28      | --        | --       | 104                               | 78.50    | 8.62      | --        | --       |
|                      | Minority | 57         | 71.24    | 11.88     | --        | --       | 57                                | 70.36    | 11.18     | --        | --       |
| Cut Score Set to 80% | White    | 56         | 87.11    | 4.20      | .46       | .18      | 49                                | 86.21    | 3.45      | .52       | .61      |
|                      | Minority | 14         | 86.37    | 3.63      |           |          | 14                                | 84.14    | 3.34      |           |          |
| Cut Score Set to 75% | White    | 69         | 85.37    | 5.27      | .58       | .42      | 66                                | 83.94    | 4.90      | .61       | .51      |
|                      | Minority | 22         | 83.18    | 5.22      |           |          | 22                                | 81.48    | 4.49      |           |          |
| Cut Score Set to 70% | White    | 86         | 82.87    | 6.97      | .68       | .45      | 84                                | 81.58    | 6.32      | .61       | .31      |
|                      | Minority | 32         | 79.78    | 6.78      |           |          | 28                                | 79.66    | 5.34      |           |          |
| Cut Score Set to 65% | White    | 95         | 81.43    | 8.00      | .81       | .59      | 96                                | 79.86    | 7.49      | .82       | .59      |
|                      | Minority | 42         | 76.71    | 8.12      |           |          | 43                                | 75.40    | 7.34      |           |          |

**Note:** Eight (8) items were removed prior to rescoring and converting each candidate's score to a percentile.

N = 161

# Case Study 4

## SD Differences Presented at Cut Scores



# Case Study 5

|                      |          | 100% Items |          |           |           |          | After Appealed Items Are Removed* |          |           |           |          |
|----------------------|----------|------------|----------|-----------|-----------|----------|-----------------------------------|----------|-----------|-----------|----------|
|                      |          | <i>N</i>   | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> | <i>N</i>                          | <i>M</i> | <i>SD</i> | <i>AI</i> | <i>d</i> |
| No Cut Score         | White    | 101        | 84.71    | 6.48      | --        | --       | 101                               | 83.40    | 6.90      | --        | --       |
|                      | Minority | 59         | 81.64    | 7.01      | --        | --       | 59                                | 80.28    | 6.90      | --        | --       |
| Cut Score Set to 80% | White    | 80         | 87.21    | 4.01      | .86       | .43      | 73                                | 86.74    | 3.93      | .80       | .44      |
|                      | Minority | 40         | 85.55    | 3.55      |           |          | 34                                | 85.10    | 3.13      |           |          |
| Cut Score Set to 75% | White    | 95         | 85.69    | 5.11      | .90       | .37      | 89                                | 85.11    | 5.01      | .90       | .45      |
|                      | Minority | 50         | 83.84    | 4.71      |           |          | 47                                | 82.96    | 4.46      |           |          |
| Cut Score Set to 70% | White    | 99         | 85.16    | 5.65      | .93       | .38      | 100                               | 83.70    | 6.23      | .89       | .27      |
|                      | Minority | 54         | 83.07    | 5.31      |           |          | 52                                | 82.12    | 4.97      |           |          |
| Cut Score Set to 65% | White    | 100        | 84.99    | 5.89      | .98       | .45      | 100                               | 83.70    | 6.23      | .98       | .46      |
|                      | Minority | 57         | 82.31    | 6.10      |           |          | 57                                | 80.83    | 6.34      |           |          |

**Note:** Five (5) items were removed prior to rescoring and converting each candidate's score to a percentile.

N = 160

# Case Study 5

## SD Differences Presented at Cut Scores

