



# **Adverse Impact of the ASVAB and Special Tests**

**Findings from the FY2023 Applicant Sample**

Nicholas Howald

*Human Resources Research Organization*

Briefing presented to the DACMPT

January 23, 2025

## Agenda

- Adverse Impact Background
- AI Analysis Findings
- Conclusions

---

# Adverse Impact Background

## What Is Adverse Impact?

- Adverse impact (AI) is the unintended discrimination of a protected class that is the result of a selection procedure (Uniform Guidelines, 1978).
- AI is **not** a property of a test. However, AI may occur when a test's scores are used as the basis for selection.
- A selection test may potentially demonstrate AI when it shows sizable mean test score differences between a majority group and a protected class (minority).
- Effect sizes of the standardized mean difference give us an index to examine a test's potential for AI.

## Fairness and Adverse Impact

- Adverse impact does **not** mean a test is biased
- Evidence for validity and fairness of the Armed Services Vocational Aptitude Battery (ASVAB):
  - There is extensive evidence supporting the validity of AFQT scores in selection (Thacker et al., 2020)
  - A study by Putka et al. (2022) using five years of applicant data showed a lack of differential prediction for the AFQT in the vast majority of analyses
  - Item writers are given sensitivity and bias guidelines, and multiple HumRRO and DTAC editors review for these factors (Harber & Day, 2023)
  - Items are pretested and any that are flagged for differential item functioning (DIF) are reviewed by experts for evidence of bias; if biased content is found, the item is not used operationally (Reeder, 2023)

## How Is Adverse Impact Assessed?

- The **four-fifths rule** is often used to determine the occurrence of AI:

“A selection rate for any race, sex, or ethnic group, which is less than four-fifths (80%) of the rate for the group with the highest rate, will generally be regarded by the Federal enforcement agencies as evidence of adverse impact.” [Section 60-3, Uniform Guidelines on Employee Selection Procedures (1978); 43 FR 38295 (August 25, 1978)]

- The **ratio comparing the selection rates** is called the ***impact ratio (IR)***:

$$IR = \frac{SR_{Focal}}{SR_{Reference}}, \text{ where SR is the selection ratio}$$

- Ideally,  $IR = 1$ , but the four-fifths rule leaves wiggle room

## How Is Adverse Impact Assessed?

- Statistical significance of the IR can be computed, as well as confidence intervals around the IR (Morris & Lobsenz, 2000):

$$Z_{IR} = \frac{\ln \frac{SR_{Foc}}{SR_{Ref}}}{\sqrt{\frac{1-SR_{Tot}}{SR_{Tot}} \left( \frac{1}{N_{Foc}} + \frac{1}{N_{Ref}} \right)}}, \text{ where SR = selection rate}$$

- $Z_{IR}$  is significant at  $\alpha = .05$  if  $|Z| > 1.96$
- Confidence interval =  $e^{(\ln(IR) \pm 1.96 SE_{IR})}$ , where

$$SE_{IR} = \sqrt{\frac{1-SR_{Foc}}{N_{Foc} SR_{Foc}} + \frac{1-SR_{Ref}}{N_{Ref} SR_{Ref}}}$$

## Adverse Impact Analyses for the ASVAB

- The four-fifths rule (80%) and accompanying statistics are applied to the Armed Forces Qualification Test (AFQT) by comparing qualification rates across the focal and reference groups of interest regarding:
  - Examinees who qualify for entry into the military (i.e., those scoring in AFQT category IIIB or higher,  $AFQT \geq 31$ )
  - Examinees who qualify for enlistment incentives (i.e., those scoring in AFQT category IIIA or higher,  $AFQT \geq 50$ )
  - AI is assessed using **initial test scores *only***
  - Significance testing is not necessarily useful in analyses with very large numbers of applicants (i.e.,  $> 2,000$ )
- How should we assess AI for individual ASVAB and Special Tests, where no direct selection occurs?



## Potential for Adverse Impact

- **Effect sizes (ES)** — standardized mean differences, commonly Cohen's  $d$ 
  - ES can be plotted and classified with respect to Cohen's (1988) standards of evaluation  
**Small**  $\geq 0.20$ ; **Moderate**  $\geq 0.50$ ; **Large**  $\geq 0.80$
- Effect sizes are computed for all group comparisons as:

$$ES = \frac{\mu_{Reference} - \mu_{Focal}}{\sigma_p}$$

where:

$\mu_{Reference}$  is the mean score in the Reference (Majority) group.

$\mu_{Focal}$  is the mean score in the Focal (Minority) group

$\sigma_p$  is the pooled standard deviation across the two groups

**Note: Positive values indicate the impact favors the majority group (i.e., the minority group is impacted negatively).**

## Confidence Intervals around Effect Sizes

- A **95% confidence interval** ( $\delta_L, \delta_U$ ) for the effect size (ES) is computed as (Hedges & Olkin, 1985):

$$\delta_L = ES - 1.96\hat{\sigma}(ES) \quad \delta_U = ES + 1.96\hat{\sigma}(ES)$$

where:

$$\hat{\sigma}(ES) = \sqrt{\frac{n_R + n_F}{n_R n_F} + \frac{ES^2}{2(n_R + n_F)}}$$

- Confidence intervals provide a boundary around an ES point estimate
  - Small boundaries indicate a more precise ES estimate
  - Large boundaries indicate a more variable ES estimate

---

# FY2023 AI Analyses Findings

## ASVAB Tests and Special Tests on ASVAB Platform

- ASVAB:** Multiple-aptitude battery that measures developed abilities and helps predict future academic and occupational success in the military (all Services).
- Special Tests:** Not part of the ASVAB but delivered on the ASVAB platform, developed to inform Service-specific classification efforts

### ASVAB Tests

AFQT			
Verbal	Math	Science/Technical	Spatial
Paragraph Comprehension (PC)	Arithmetic Reasoning (AR)	General Science (GS)	Assembling Objects (AO)
Word Knowledge (WK)	Math Knowledge (MK)	Electronics Information (EI)	
		Mechanical Comprehension (MC)	
		Auto Information/Shop Information (AS)	

### Special Tests

Cyber Test (CT)	Coding Speed (CS)
Test of basic computer and information systems knowledge (all Services)	A speeded test of assigning code numbers to words (Navy only)

## Current ASVAB AI Analyses

- Sample: FY2023 applicants
- Tests: ASVAB AFQT (IIIA+ and IIIB+), ASVAB Subtests, Cyber Test, and Coding Speed
- Group comparisons:

Pair	Reference Group	Focal Group
1	Males	Females
2	Non-Hispanic Whites	Hispanic Whites
3	Non-Hispanic Whites	Non-Hispanic Blacks
4	Non-Hispanic Whites	Non-Hispanic Asians
5	Non-Hispanic Whites	Non-White Hispanics

- The focal group is potentially disadvantaged relative to the reference group.
- All included groups represent > 2% of the applicant population.

# Current ASVAB AI Analyses

## Data Cleaning

- Initial test record with valid score, name, and SSN
  - **ASVAB:** 248,434 (Total); 163,132 (CAT); 1,003 (P&P); 84,299 (Verified PiCAT)
  - **CT:** 60,230
  - **CS:** 41,145
- Remove duplicates across assessments (ASVAB ONLY): 247,779 (*n* = 655 removed)
- Timing: >2.5 SD below mean response time
  - **ASVAB:** 247,754 (*n* = 25 removed); **CT:** 60,210 (*n* = 20 removed)
- Timing: < 2 minutes to complete assessment
  - **CS:** 40,984 (*n* = 161 removed)
- Missing on all demographic variables (i.e., gender, race, and ethnicity)
  - **ASVAB:** 241,412 (*n* = 6,342 removed)
  - **CT:** 49,681 (*n* = 10,529 removed)
  - **CS:** 39,213 (*n* = 1,771 removed)

## Current ASVAB AI Analyses

Sample Category	ASVAB N	ASVAB Percent	Cyber Test N	Cyber Test Percent	Coding Speed N	Coding Speed Percent
Males	183,108	76%	38,710	78%	28,597	73%
Females	58,304	24%	10,971	22%	10,616	27%
Non-Hispanic Whites	92,151	38%	19,492	39%	13,335	34%
Hispanic Whites	58,622	24%	12,182	25%	8,459	22%
Non-Hispanic Blacks	62,416	26%	10,580	21%	10,693	27%
Non-Hispanic Asians	12,836	5%	2,882	6%	2,702	7%
Non-White Hispanics	6,553	3%	1,846	4%	1,623	4%
<b>Total</b>	<b>241,412</b>	<b>-</b>	<b>49,681</b>	<b>-</b>	<b>39,213</b>	<b>-</b>

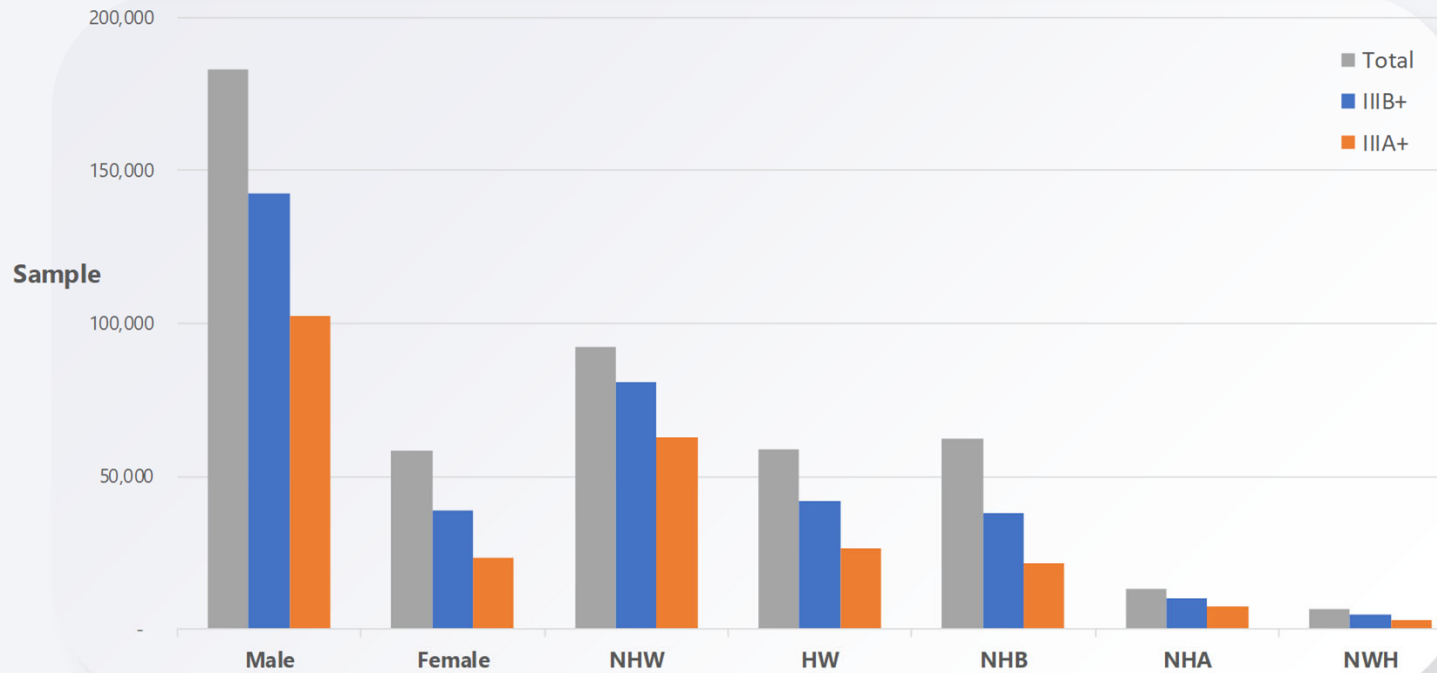
*Note.* Ethnicity does not add up to 100% due to missing data or other sample category values below the 2% threshold for some individuals. Some individuals in CT ( $n = 2,221$ ; 4%) and CS ( $n = 5,193$ ; 13%) did not have corresponding data in the ASVAB sample, as they had taken ASVAB during FY22 (Oct 1, 2021 – Sep 30, 2022).

---

# Adverse Impact Analysis Findings



## Adverse Impact Analysis Sample Sizes for FY23



NHW: Non-Hispanic White  
 HW: Hispanic White  
 NHB: Non-Hispanic Black  
 NHA: Non-Hispanic Asian  
 NWH: Non-White Hispanic

---

# Impact Ratios for IIB+ and IIIA+

# Impact Ratios for AFQT Cutscores FY2023 IIB+ and IIIA+



## Adverse Impact Ratios

Threshold

IIIA+

IIB+

Comparison	Impact Ratio
Female vs Male	0.85
HW vs NHW	0.81
NHB vs NHW	0.69
NHA vs NHW	0.88
NWH vs NHW	0.83

Trivial Small Mod Large

NHW: Non-Hispanic White  
 HW: Hispanic White  
 NHB: Non-Hispanic Black  
 NHA: Non-Hispanic Asian  
 NWH: Non-White Hispanic

# Impact Ratios for AFQT Cutscores FY2023 IIB+ and IIIA+



## Adverse Impact Ratios

Threshold

IIB+

IIIB+

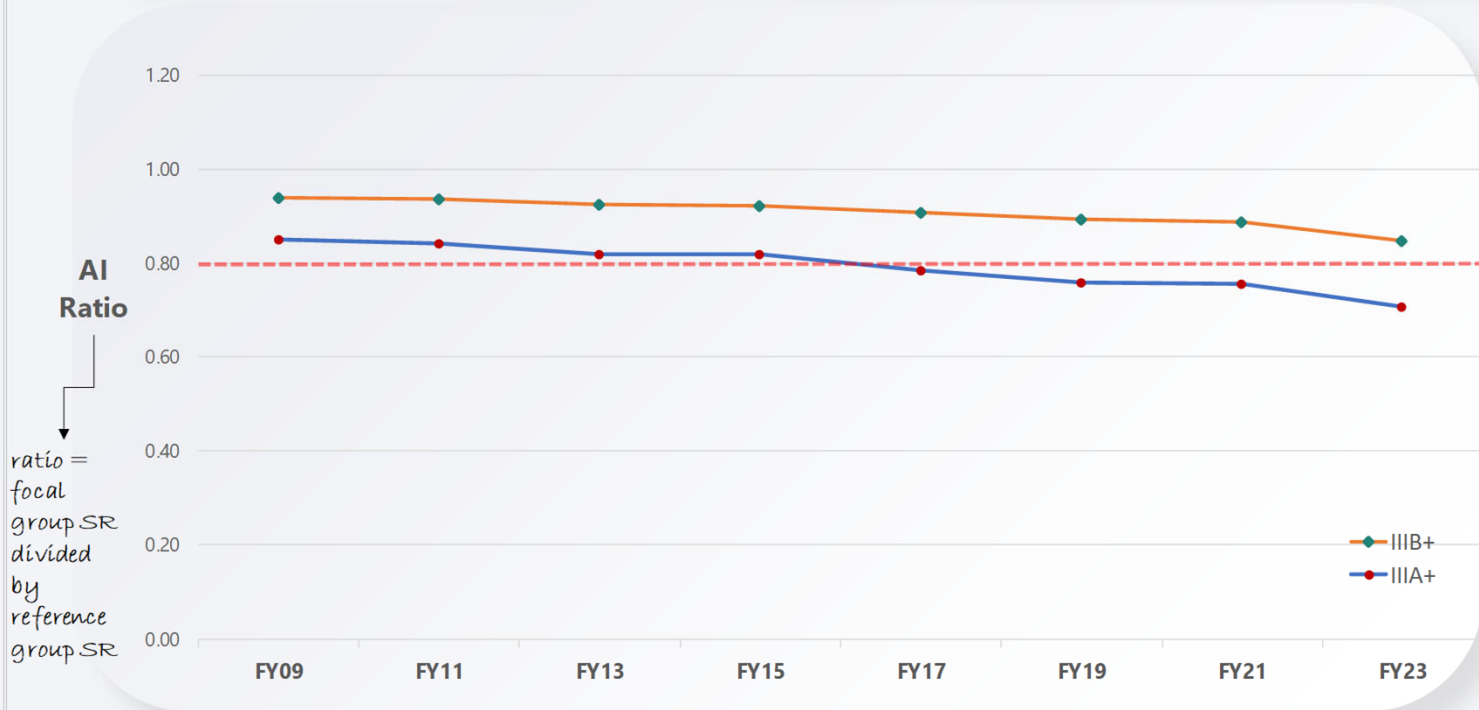
### Comparison Impact Ratio

Female vs Male	0.71
HW vs NHW	0.66
NHB vs NHW	0.50
NHA vs NHW	0.86
NWH vs NHW	0.69

Trivial Small Mod Large

NHW: Non-Hispanic White  
 HW: Hispanic White  
 NHB: Non-Hispanic Black  
 NHA: Non-Hispanic Asian  
 NWH: Non-White Hispanic

# Comparison of Impact Ratios for Odd-Numbered FYs 09-23



## Adverse Impact Ratios

Comparison

**Female vs Male**

HW vs NHW

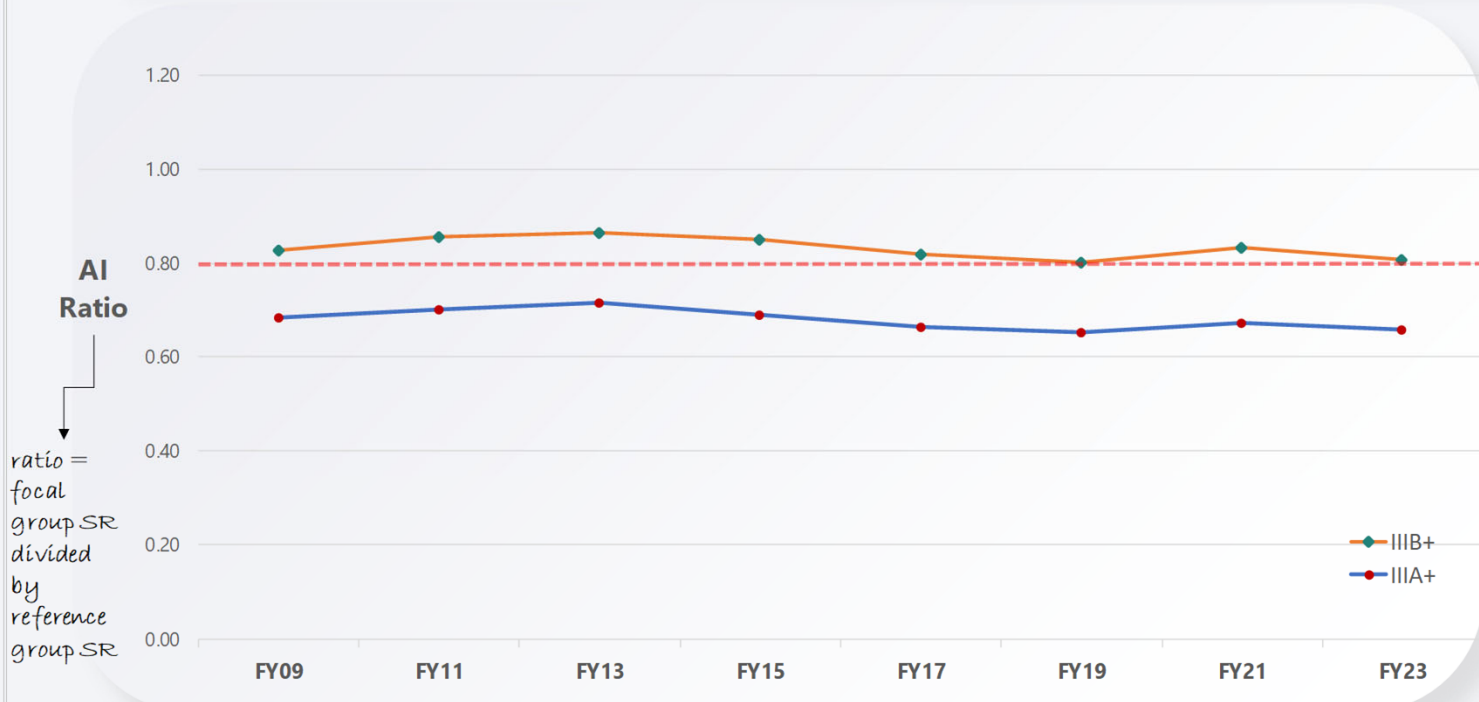
NHA vs NHW

NHB vs NHW

Year	III B+	III A+
FY09	<b>0.94</b>	<b>0.85</b>
FY11	<b>0.94</b>	<b>0.84</b>
FY13	<b>0.93</b>	<b>0.82</b>
FY15	<b>0.92</b>	<b>0.82</b>
FY17	<b>0.91</b>	<b>0.79</b>
FY19	<b>0.89</b>	<b>0.76</b>
FY21	<b>0.89</b>	<b>0.75</b>
FY23	<b>0.85</b>	<b>0.71</b>

Trivial Small Mod Large

# Comparison of Impact Ratios for Odd-Numbered FYs 09-23



## Adverse Impact Ratios

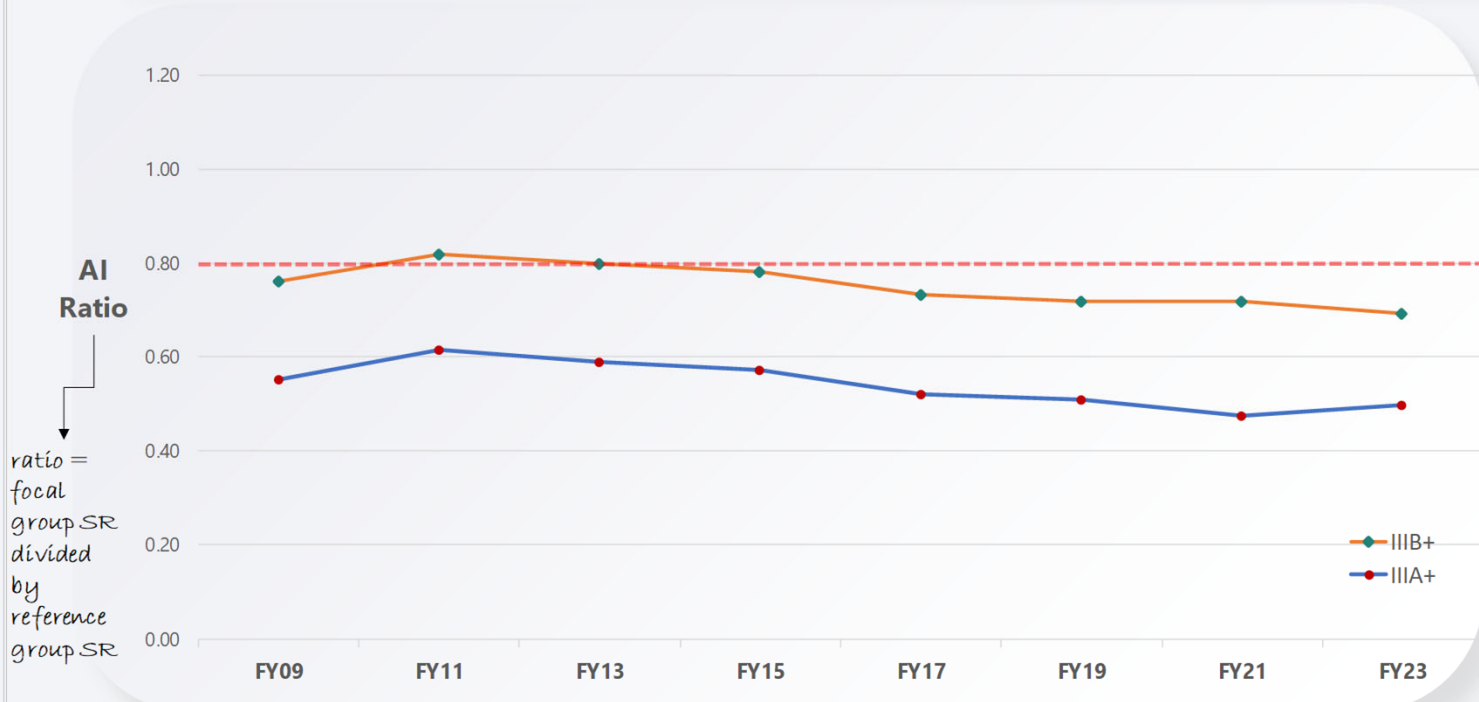
Comparison

Female vs Male	<b>HW vs NHW</b>
NHA vs NHW	NHB vs NHW

Year	III B+	III A+
FY09	<b>0.83</b>	<b>0.68</b>
FY11	<b>0.86</b>	<b>0.70</b>
FY13	<b>0.87</b>	<b>0.72</b>
FY15	<b>0.85</b>	<b>0.69</b>
FY17	<b>0.82</b>	<b>0.66</b>
FY19	<b>0.80</b>	<b>0.65</b>
FY21	<b>0.83</b>	<b>0.67</b>
FY23	<b>0.81</b>	<b>0.66</b>

Trivial Small Mod Large

# Comparison of Impact Ratios for Odd-Numbered FYs 09-23



## Adverse Impact Ratios

Comparison

Female vs Male

HW vs NHW

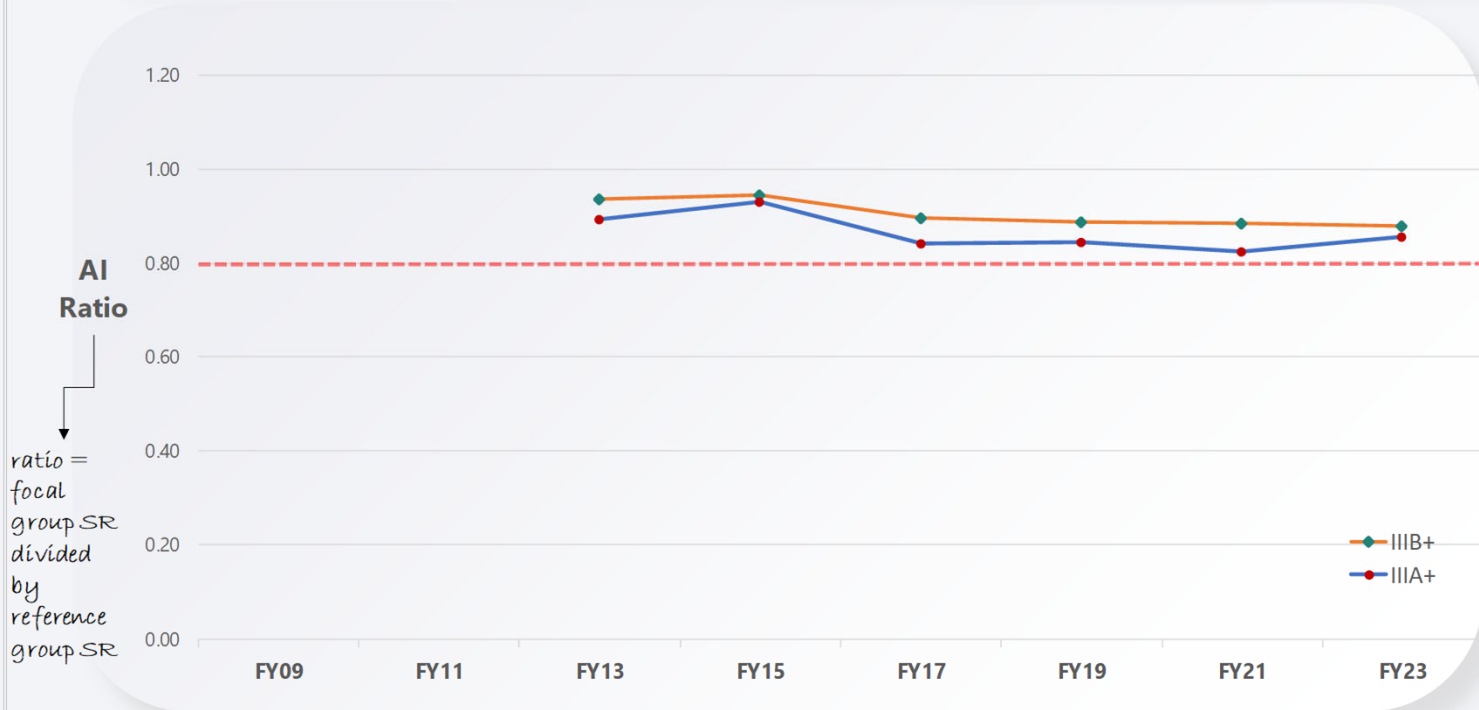
NHA vs NHW

**NHB vs NHW**

Year	IIIB+	IIIA+
FY09	<b>0.76</b>	<b>0.55</b>
FY11	<b>0.82</b>	<b>0.62</b>
FY13	<b>0.80</b>	<b>0.59</b>
FY15	<b>0.78</b>	<b>0.57</b>
FY17	<b>0.73</b>	<b>0.52</b>
FY19	<b>0.72</b>	<b>0.51</b>
FY21	<b>0.72</b>	<b>0.47</b>
FY23	<b>0.69</b>	<b>0.50</b>

Trivial Small Mod Large

# Comparison of Impact Ratios for Odd-Numbered FYs 09-23



AI Ratio  
 $\text{ratio} = \frac{\text{focal group SR}}{\text{reference group SR}}$

## Adverse Impact Ratios

Comparison

Female vs Male

HW vs NHW

**NHA vs NHW**

NHB vs NHW

Year	III B+	III A+
FY09		
FY11		
FY13	<b>0.93</b>	<b>0.89</b>
FY15	<b>0.95</b>	<b>0.93</b>
FY17	<b>0.90</b>	<b>0.84</b>
FY19	<b>0.89</b>	<b>0.84</b>
FY21	<b>0.88</b>	<b>0.82</b>
FY23	<b>0.88</b>	<b>0.86</b>

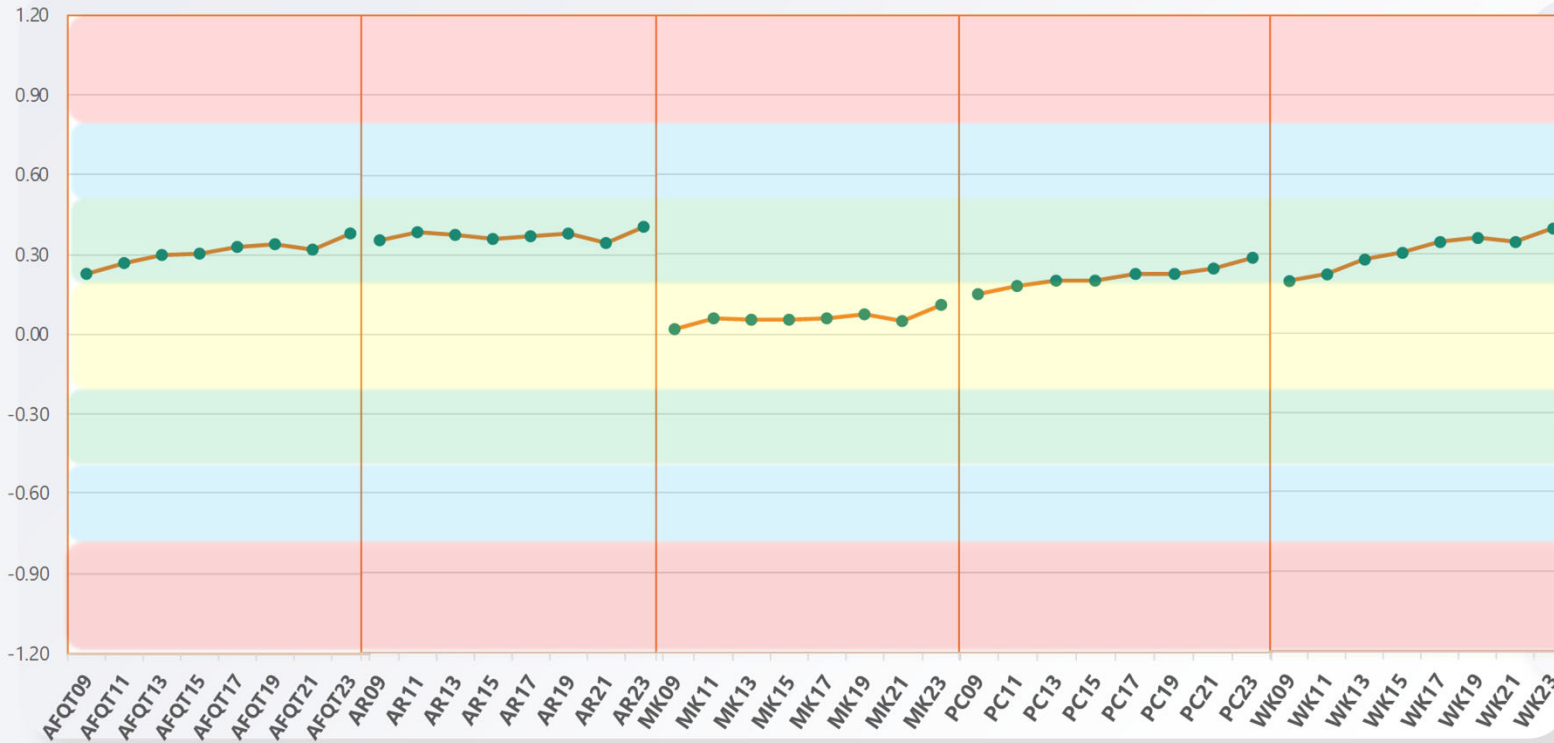
Trivial Small **Mod** Large



---

# Effect Sizes over Time

# Comparison of Effect Sizes for Odd-Numbered FY 09-23 (AFQT Tests/Scores)



### Effect Sizes

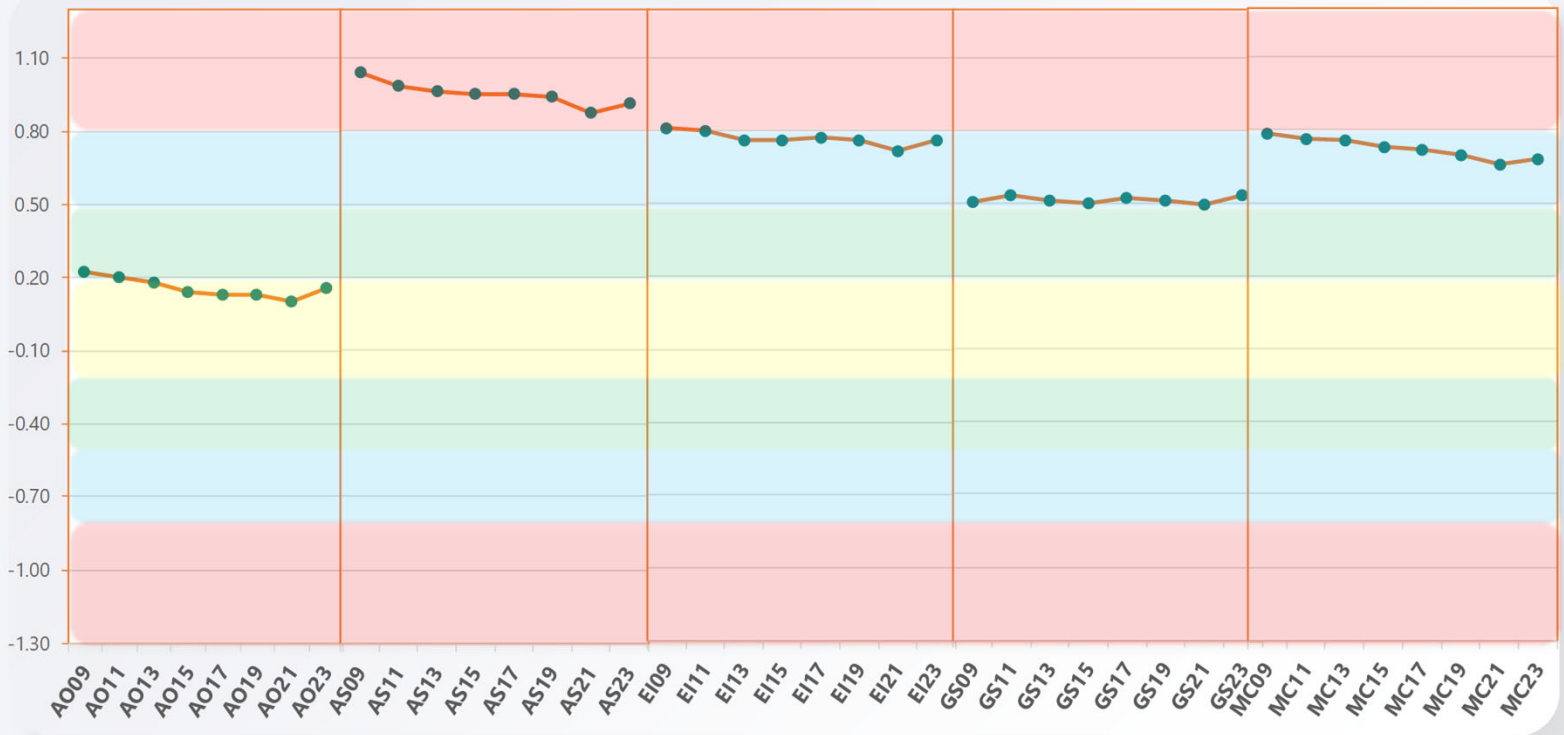
Comparison

Female vs Male     HW vs NHW  
 NHA vs NHW     NHB vs NHW

Year	AFQT	AR	MK	PC	WK
FY09	0.23	0.35	0.02	0.15	0.20
FY11	0.27	0.38	0.06	0.18	0.23
FY13	0.30	0.37	0.05	0.20	0.28
FY15	0.30	0.36	0.05	0.20	0.31
FY17	0.33	0.37	0.06	0.23	0.35
FY19	0.34	0.38	0.07	0.23	0.36
FY21	0.32	0.34	0.05	0.25	0.35
FY23	0.38	0.40	0.11	0.29	0.40

Trivial    Small    Mod    Large

# Comparison of Effect Sizes for Odd-Numbered FY 09-23 (non-AFQT Tests)



### Effect Sizes

Comparison

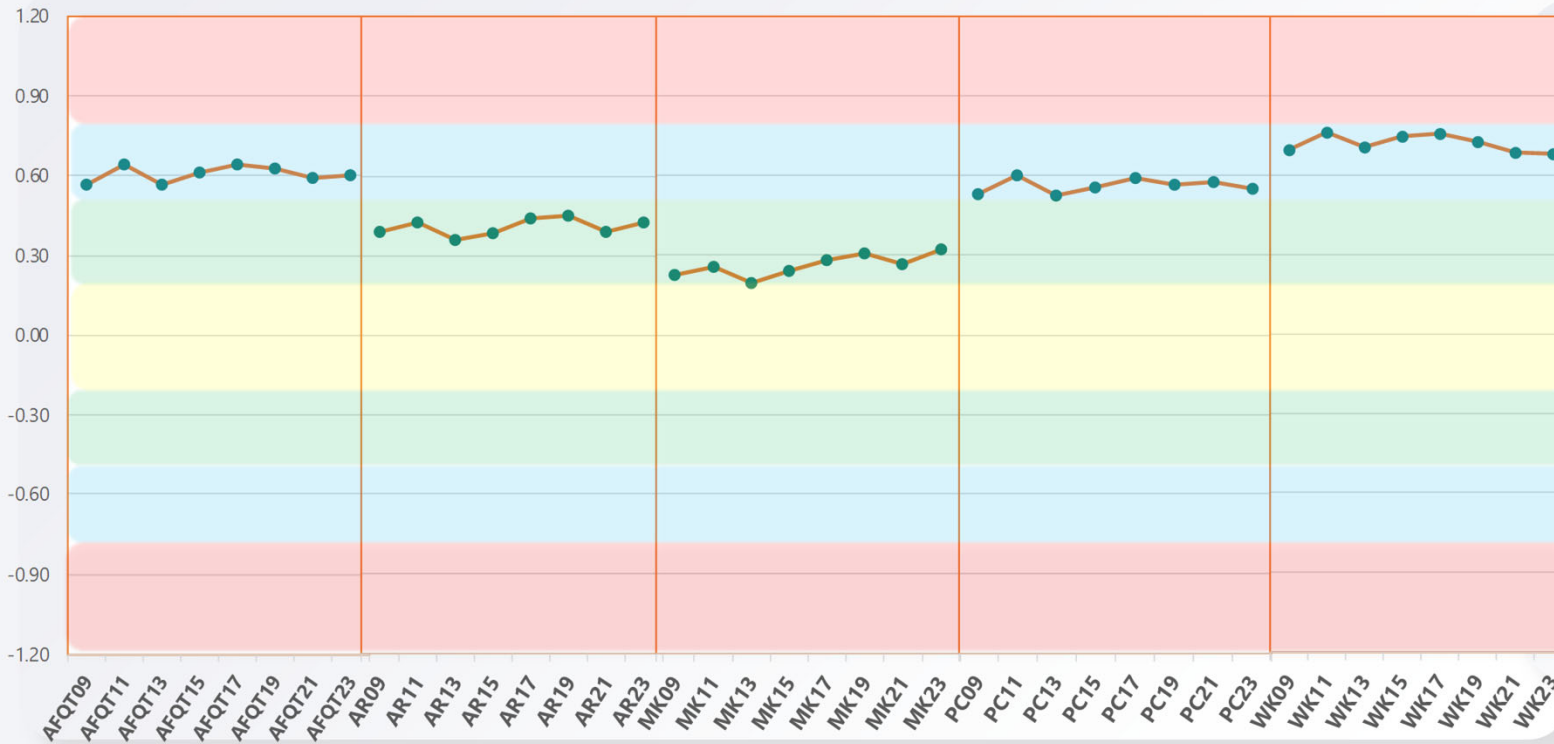
**Female vs Male**  HW vs NHW

NHA vs NHW  NHB vs NHW

Year	AO	AS	EI	GS	MC
FY09	0.22	1.04	0.81	0.51	0.79
FY11	0.20	0.98	0.80	0.53	0.77
FY13	0.18	0.96	0.76	0.51	0.76
FY15	0.14	0.95	0.76	0.50	0.73
FY17	0.13	0.95	0.77	0.53	0.72
FY19	0.13	0.94	0.76	0.51	0.70
FY21	0.10	0.87	0.72	0.50	0.66
FY23	0.16	0.91	0.76	0.53	0.68

Trivial Small Mod Large

## Comparison of Effect Sizes for Odd-Numbered FY 09-23 (AFQT Tests/Scores)



### Effect Sizes

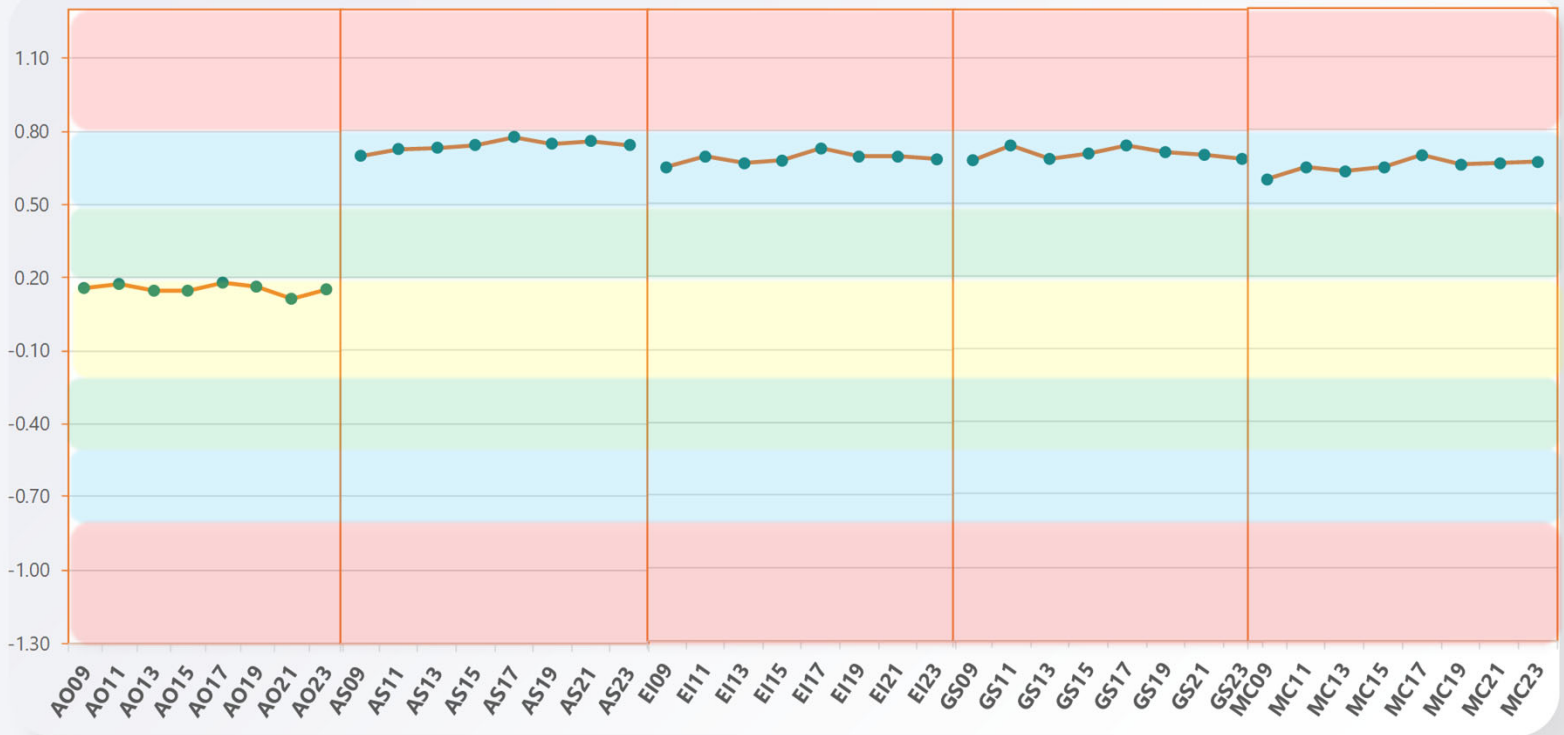
Comparison

- Female vs Male  
 HW vs NHW  
 NHA vs NHW  
 NHB vs NHW

Year	AFQT	AR	MK	PC	WK
FY09	0.57	0.39	0.23	0.53	0.70
FY11	0.64	0.42	0.25	0.60	0.76
FY13	0.57	0.36	0.20	0.53	0.71
FY15	0.61	0.39	0.24	0.56	0.75
FY17	0.64	0.44	0.28	0.59	0.76
FY19	0.63	0.45	0.30	0.57	0.73
FY21	0.59	0.39	0.26	0.58	0.69
FY23	0.60	0.42	0.32	0.55	0.68

Trivial
Small
Mod
Large

## Comparison of Effect Sizes for Odd-Numbered FY 09-23 (non-AFQT Tests)



### Effect Sizes

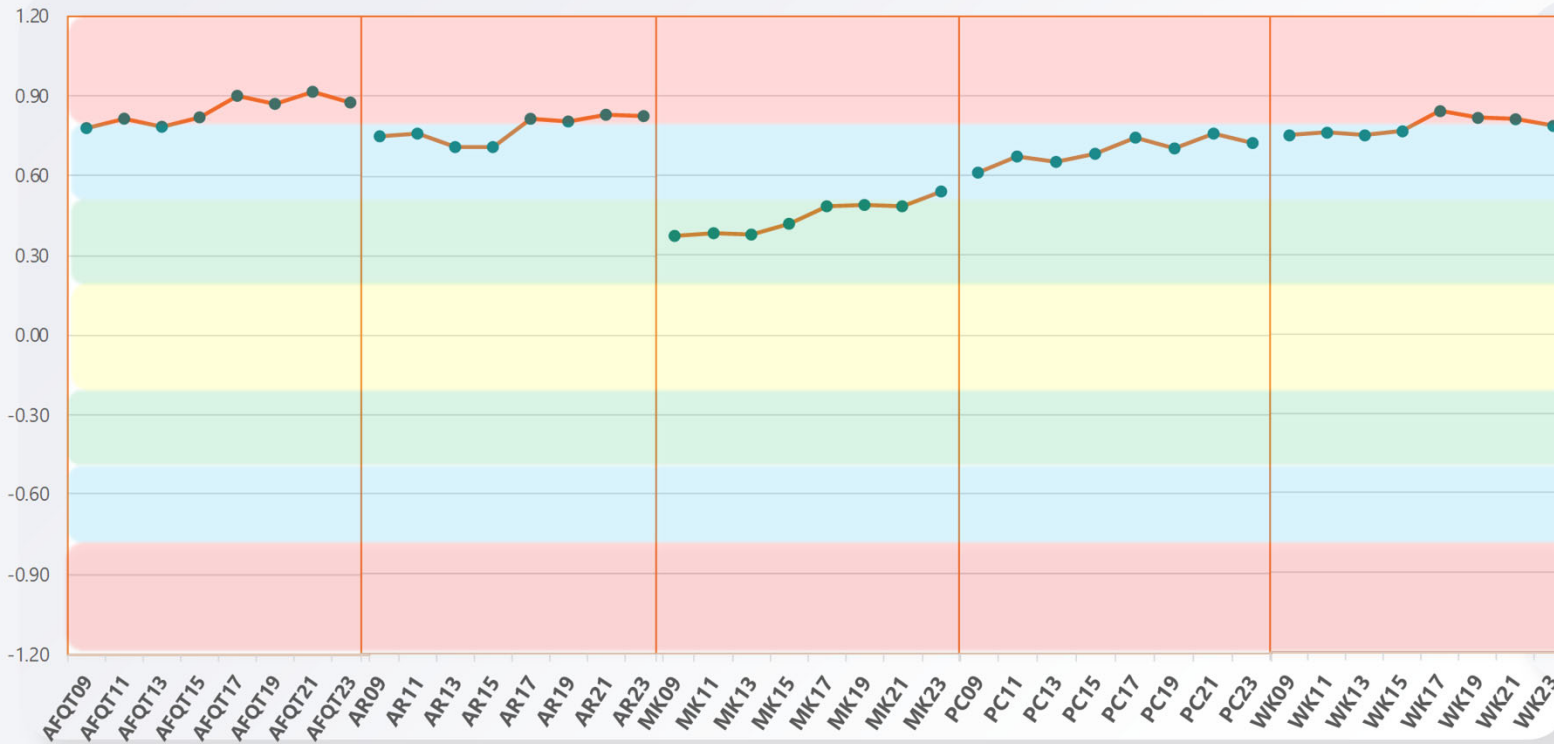
Comparison

- Female vs Male  HW vs NHW
- NHA vs NHW  NHB vs NHW

Year	AO	AS	EI	GS	MC
FY09	0.16	0.70	0.65	0.68	0.60
FY11	0.17	0.73	0.70	0.74	0.65
FY13	0.14	0.73	0.67	0.68	0.63
FY15	0.15	0.74	0.68	0.71	0.65
FY17	0.18	0.77	0.73	0.74	0.70
FY19	0.16	0.75	0.70	0.71	0.66
FY21	0.11	0.76	0.70	0.70	0.66
FY23	0.15	0.74	0.68	0.68	0.67

Trivial Small Mod Large

## Comparison of Effect Sizes for Odd-Numbered FY 09-23 (AFQT Tests/Scores)



### Effect Sizes

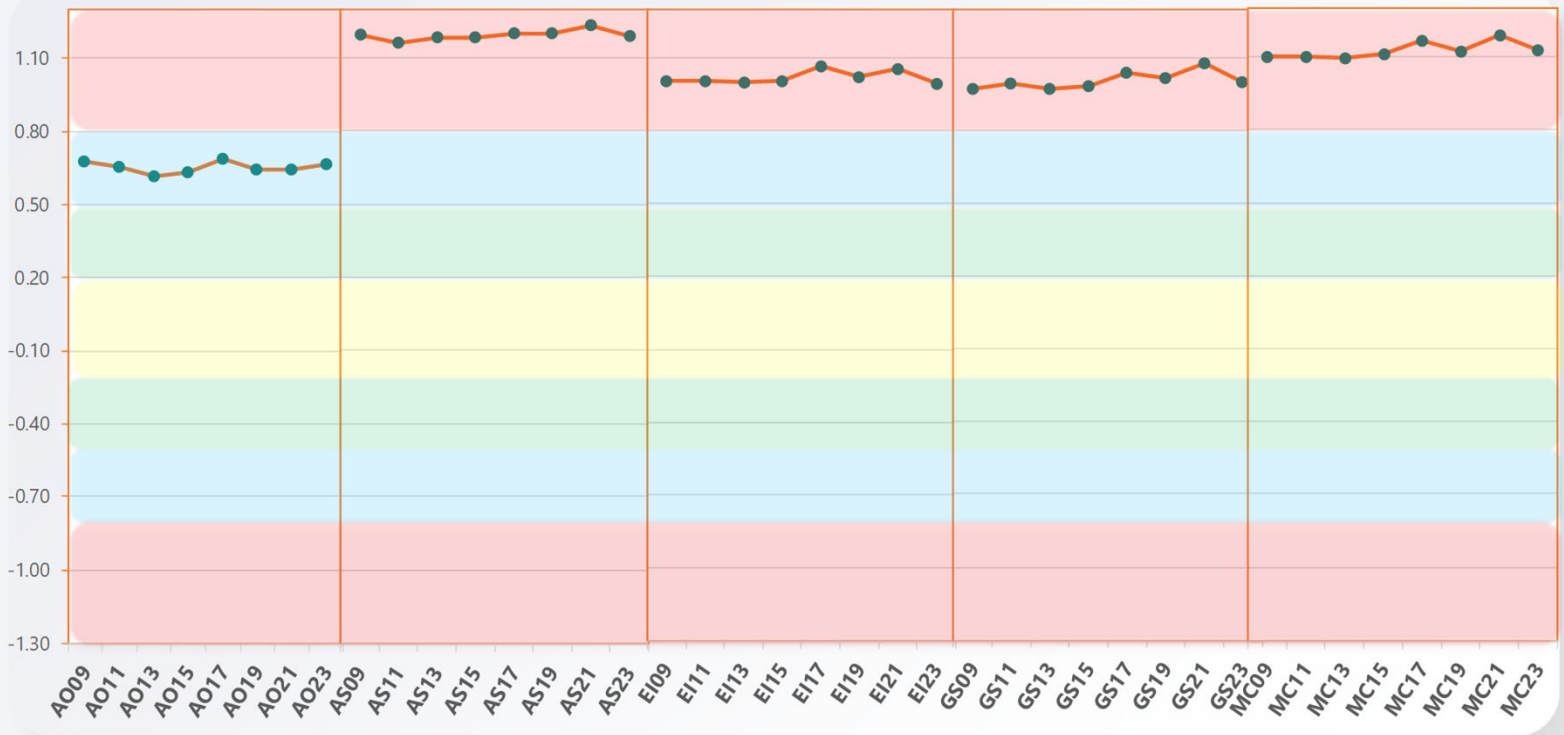
Comparison

- Female vs Male
- HW vs NHW
- NHA vs NHW
- NHB vs NHW

Year	AFQT	AR	MK	PC	WK
FY09	0.78	0.75	0.38	0.61	0.75
FY11	0.82	0.76	0.38	0.67	0.76
FY13	0.78	0.71	0.38	0.65	0.76
FY15	0.82	0.71	0.42	0.68	0.77
FY17	0.90	0.82	0.49	0.75	0.85
FY19	0.87	0.80	0.49	0.70	0.82
FY21	0.91	0.83	0.48	0.76	0.81
FY23	0.87	0.83	0.54	0.73	0.79

Trivial
Small
Mod
Large

## Comparison of Effect Sizes for Odd-Numbered FY 09-23 (non-AFQT Tests)



### Effect Sizes

Comparison

Female vs Male

HW vs NHW

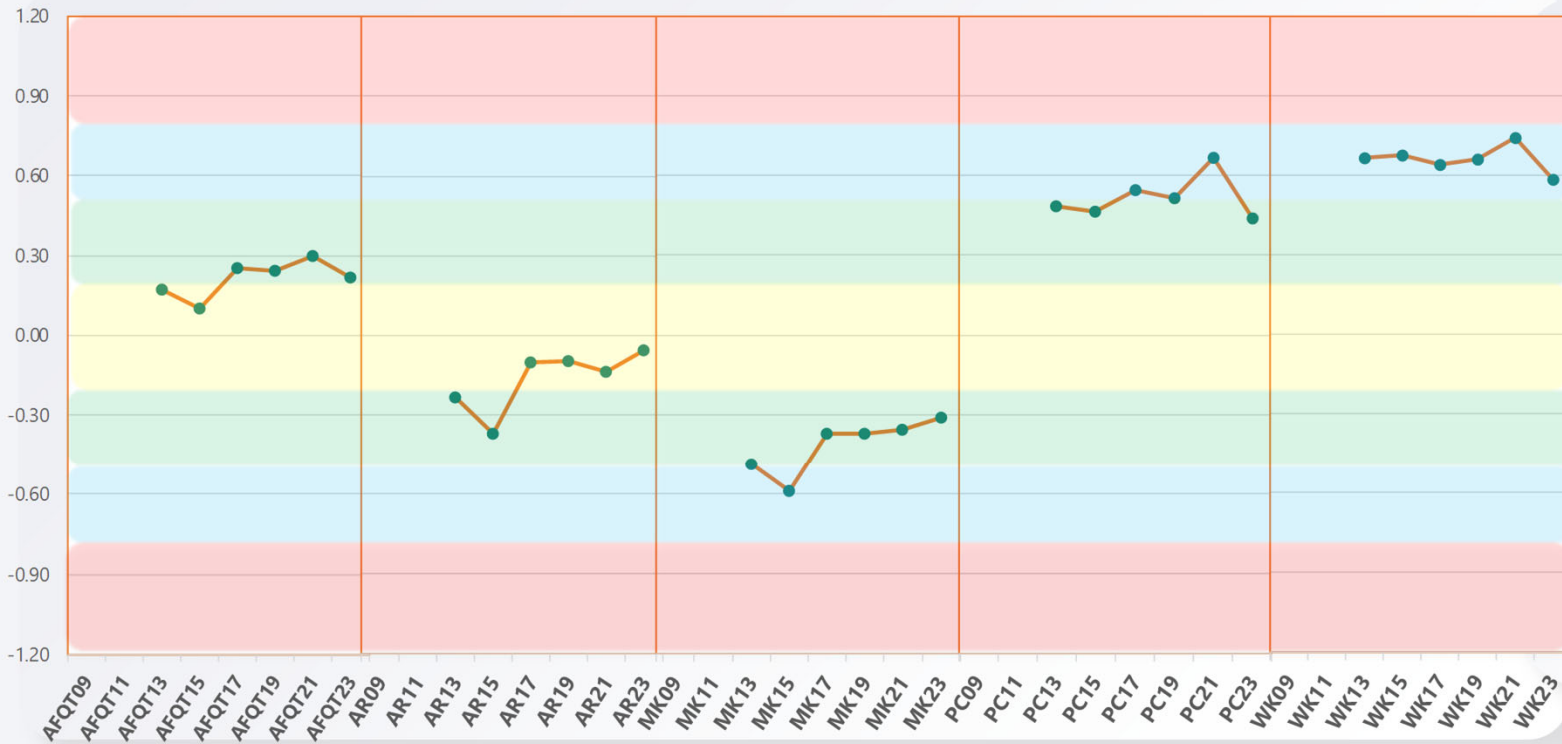
NHA vs NHW

**NHB vs NHW**

Year	AO	AS	EI	GS	MC
FY09	0.68	1.19	1.01	0.97	1.10
FY11	0.65	1.16	1.01	0.99	1.10
FY13	0.62	1.18	1.00	0.97	1.10
FY15	0.63	1.18	1.01	0.98	1.11
FY17	0.69	1.20	1.07	1.04	1.17
FY19	0.64	1.20	1.02	1.02	1.13
FY21	0.64	1.23	1.06	1.08	1.19
FY23	0.66	1.19	1.00	1.00	1.13

Trivial Small Mod Large

# Comparison of Effect Sizes for Odd-Numbered FY 09-23 (AFQT Tests/Scores)



**Effect Sizes**

Comparison

Female vs Male  HW vs NHW

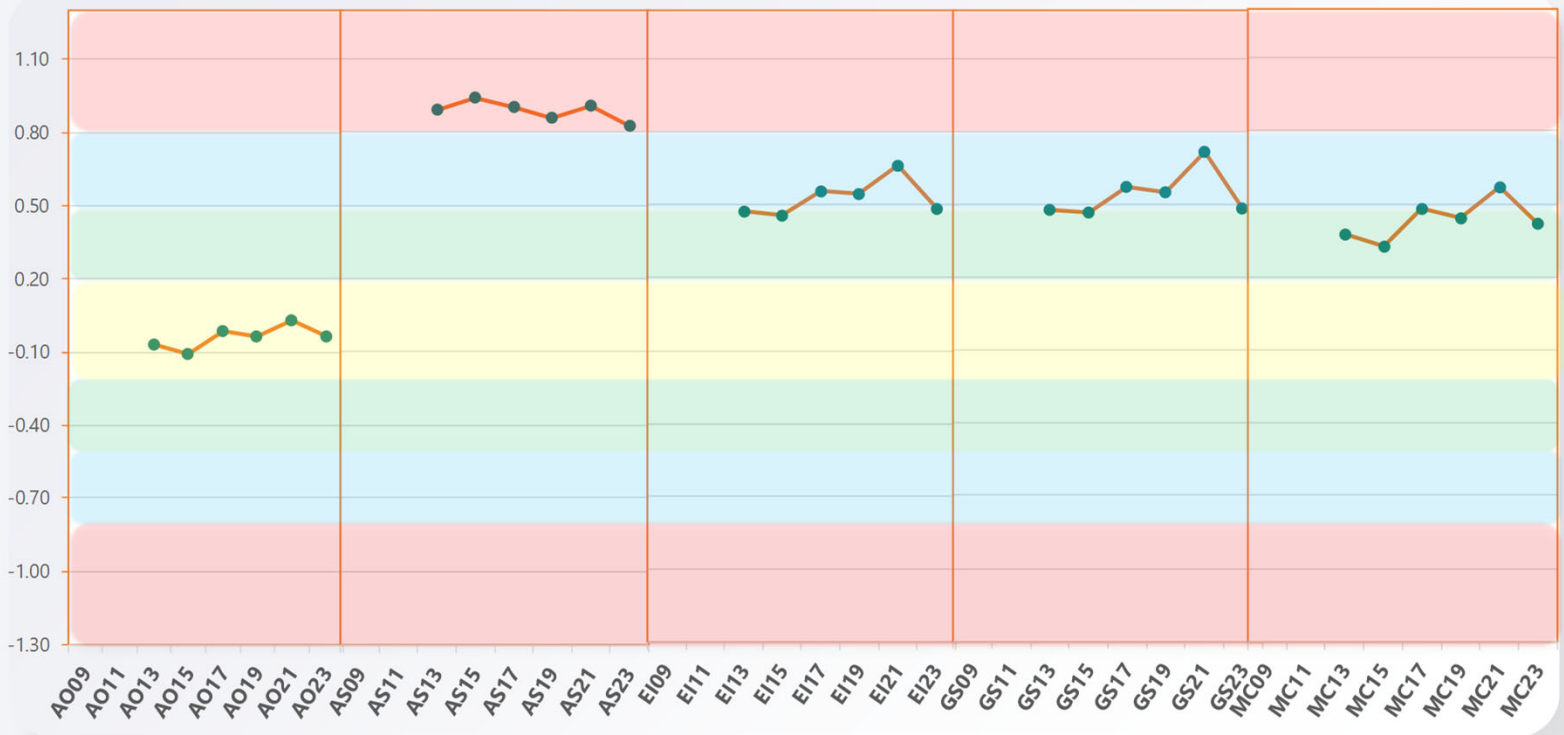
**NHA vs NHW**  NHB vs NHW

Year	AFQT	AR	MK	PC	WK
FY09					
FY11					
FY13	0.17	-0.24	-0.49	0.49	0.67
FY15	0.10	-0.37	-0.59	0.46	0.68
FY17	0.25	-0.10	-0.37	0.55	0.64
FY19	0.24	-0.10	-0.37	0.51	0.66
FY21	0.30	-0.14	-0.36	0.67	0.74
FY23	0.21	-0.06	-0.31	0.44	0.58

Trivial
Small
Mod
Large



# Comparison of Effect Sizes for Odd-Numbered FY 09-23 (non-AFQT Tests)



## Effect Sizes

Comparison

Female vs Male    HW vs NHW

**NHA vs NHW**    NHB vs NHW

Year	AO	AS	EI	GS	MC
FY09					
FY11					
FY13	-0.07	0.89	0.48	0.48	0.38
FY15	-0.11	0.94	0.46	0.47	0.33
FY17	-0.01	0.90	0.56	0.57	0.48
FY19	-0.04	0.86	0.55	0.55	0.44
FY21	0.03	0.91	0.66	0.72	0.57
FY23	-0.04	0.83	0.49	0.49	0.42

Trivial    Small    Mod    Large

---

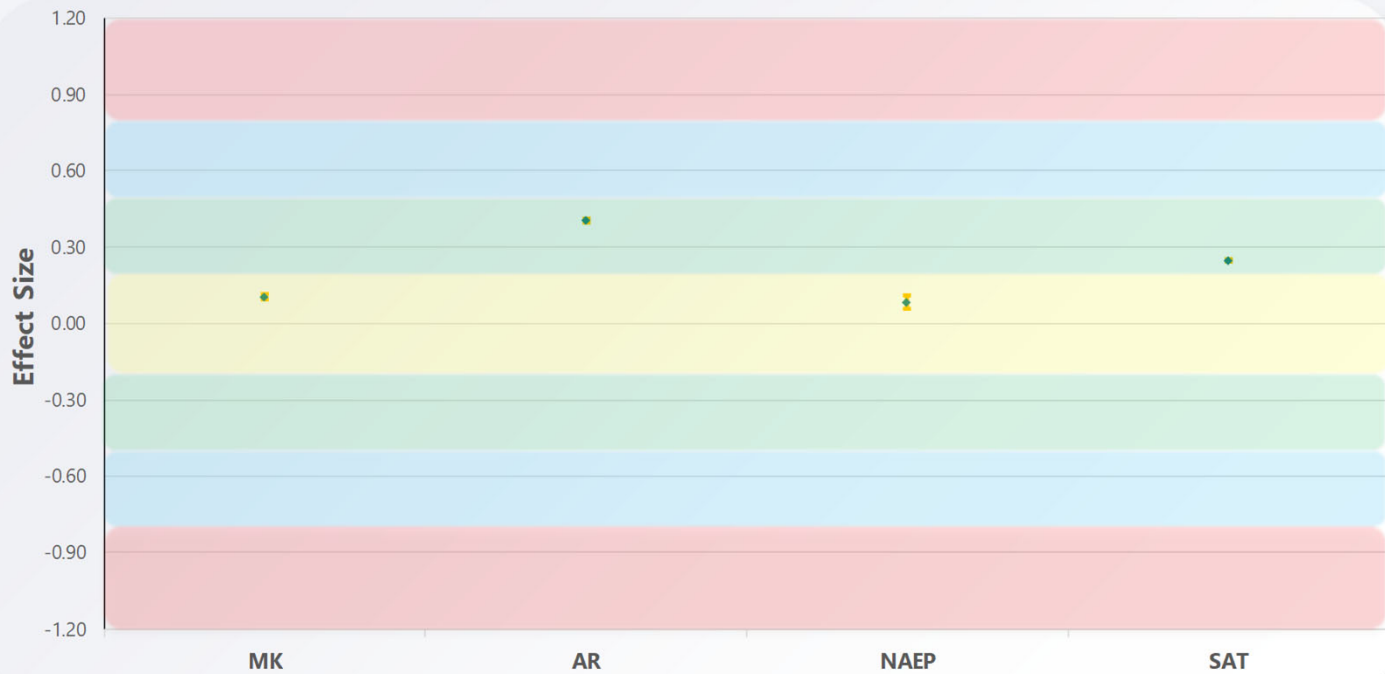
# Comparison with Other Large-Scale Testing Programs

## What Does It Mean?

- The magnitude of impact on the ASVAB has remained fairly consistent across fiscal years, but still varies in size from negligible to large across tests and groups.
- A comparison of impact across different testing programs gives some indication of whether the observed FY2023 magnitudes are reasonable.
- Sufficient information for estimating effect sizes is available online for two other large-scale testing programs:
  1. SAT\* — 2016 College-Bound Seniors (Math and Reading)
  2. NAEP — 2019 Grade 12 (Reading, Math, and Science)

\*SAT stopped reporting SDs for demographic comparisons after 2016 in publicly available online content, limiting the ability to calculate effect sizes for more recent years without submitting data requests.

# Comparison of Effect Sizes Across Testing Programs (Math)



## Effect Sizes

Comparison

**Female vs Male**

H vs NHW

NHA vs NHW

NHB vs NHW

**Test**   **Effect Size**

MK   **0.11**

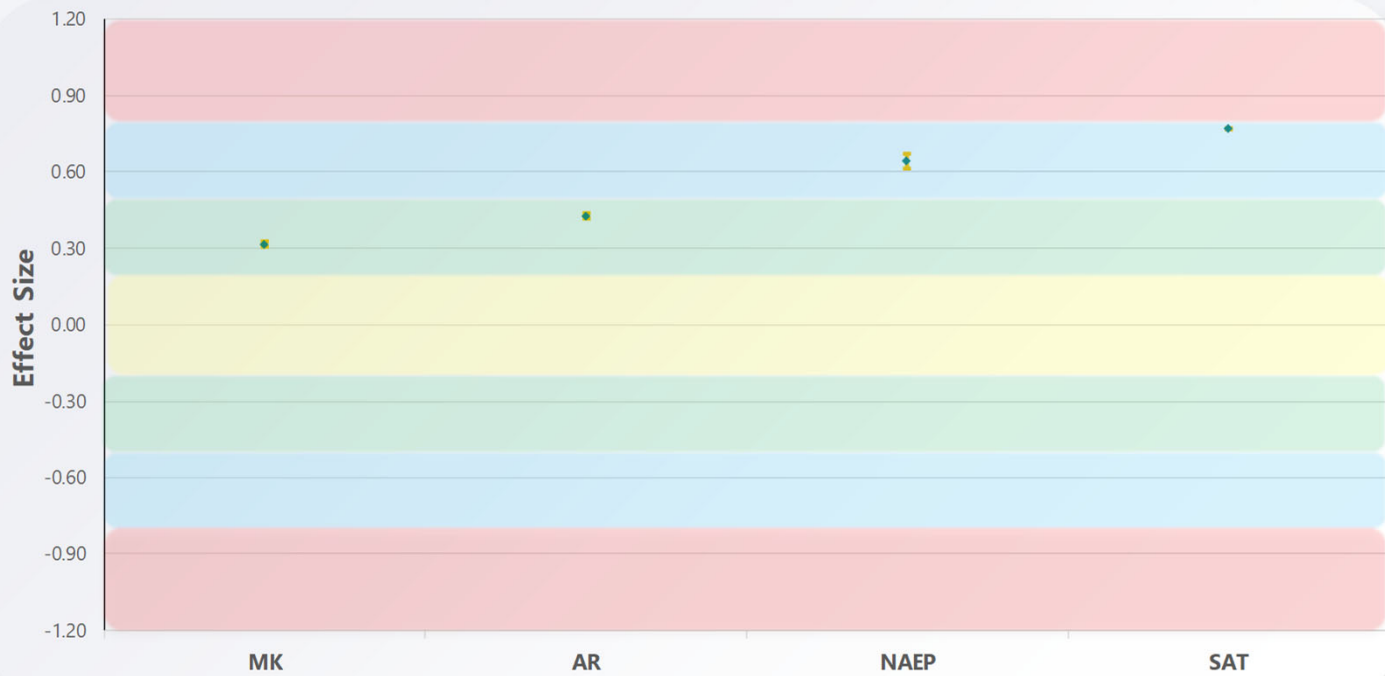
AR   **0.40**

NAEP   **0.08**

SAT   **0.25**

Trivial   Small   Mod   Large

# Comparison of Effect Sizes Across Testing Programs (Math)



## Effect Sizes

Comparison

Female vs Male

**H vs NHW**

NHA vs NHW

NHB vs NHW

Test Effect Size

MK	0.32
AR	0.43
NAEP	0.64
SAT	0.77

Trivial Small Mod Large

## Comparison of Effect Sizes Across Testing Programs (Math)



### Effect Sizes

Comparison

Female vs Male

H vs NHW

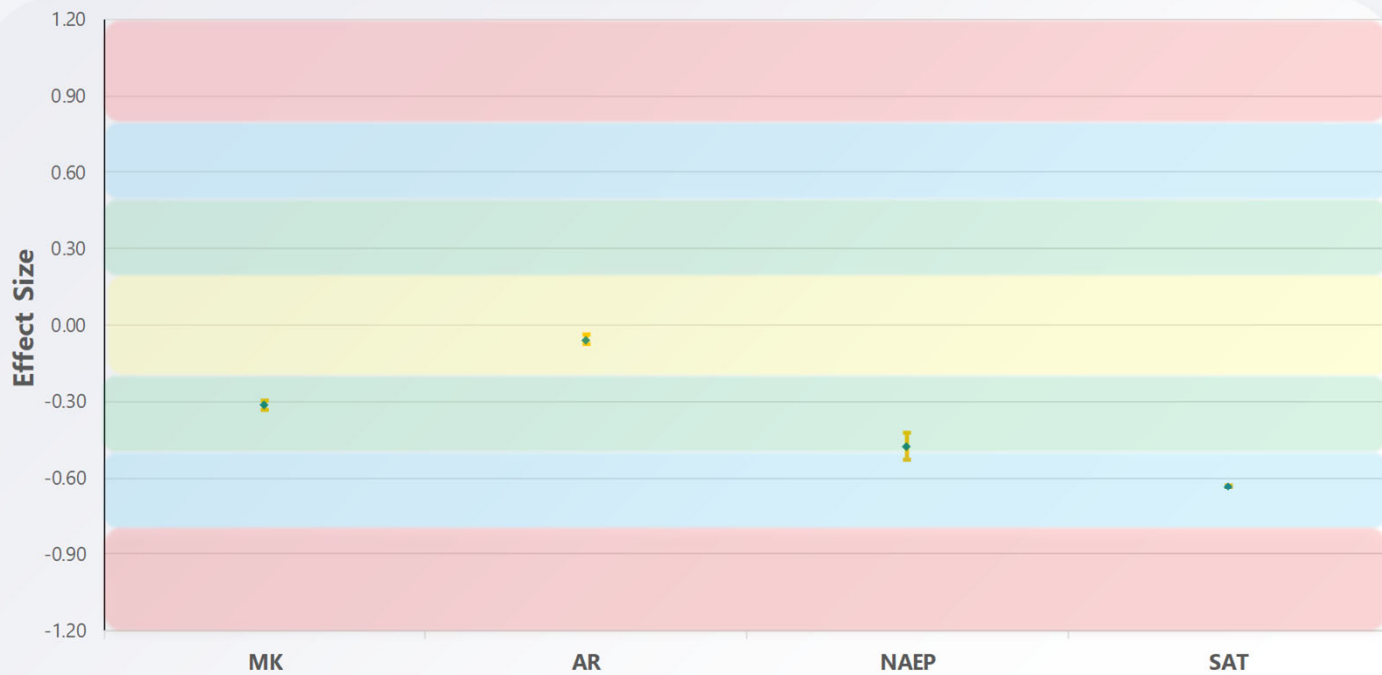
NHA vs NHW

**NHB vs NHW**

Test	Effect Size
MK	0.54
AR	0.83
NAEP	0.95
SAT	1.04

Trivial Small Mod Large

## Comparison of Effect Sizes Across Testing Programs (Math)



### Effect Sizes

Comparison

Female vs Male

H vs NHW

**NHA vs NHW**

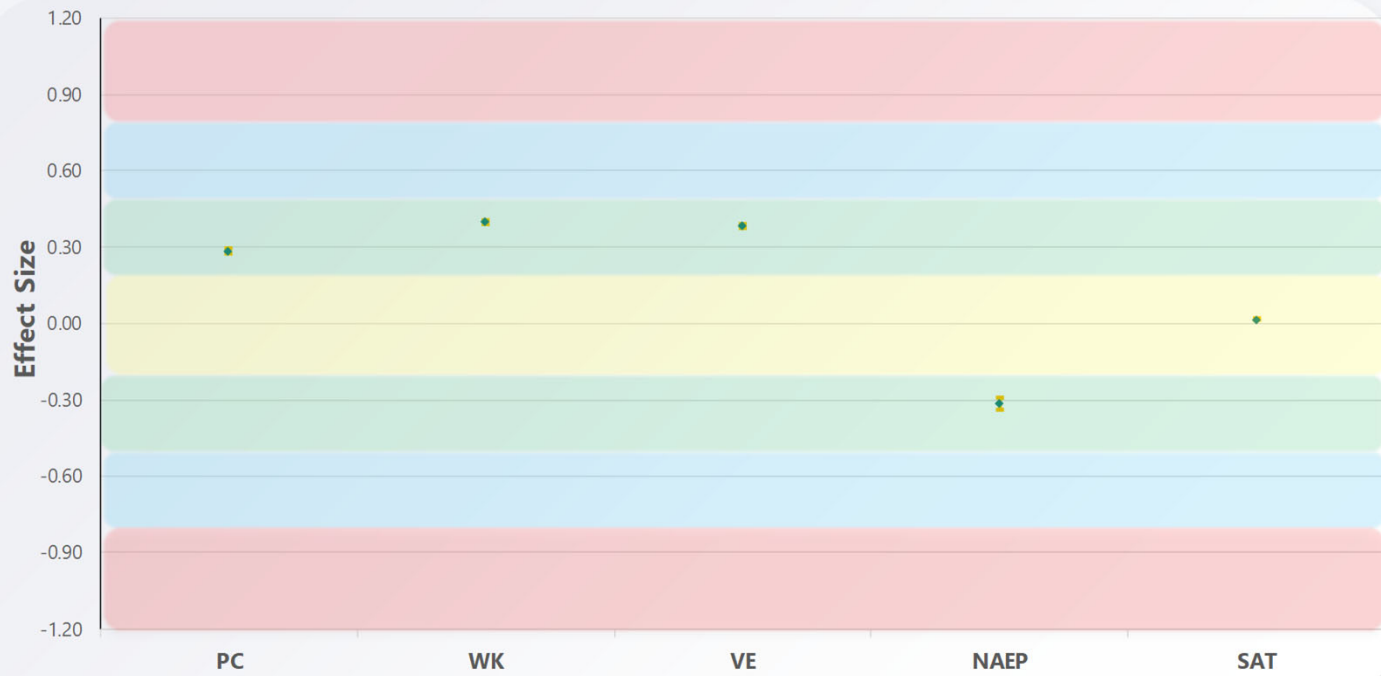
NHB vs NHW

Test Effect Size

MK	-0.31
AR	-0.06
NAEP	-0.48
SAT	-0.63

Trivial Small Mod Large

# Comparison of Effect Sizes Across Testing Programs (Reading)



## Effect Sizes

Comparison

Female vs Male

H vs NHW

NHA vs NHW

NHB vs NHW

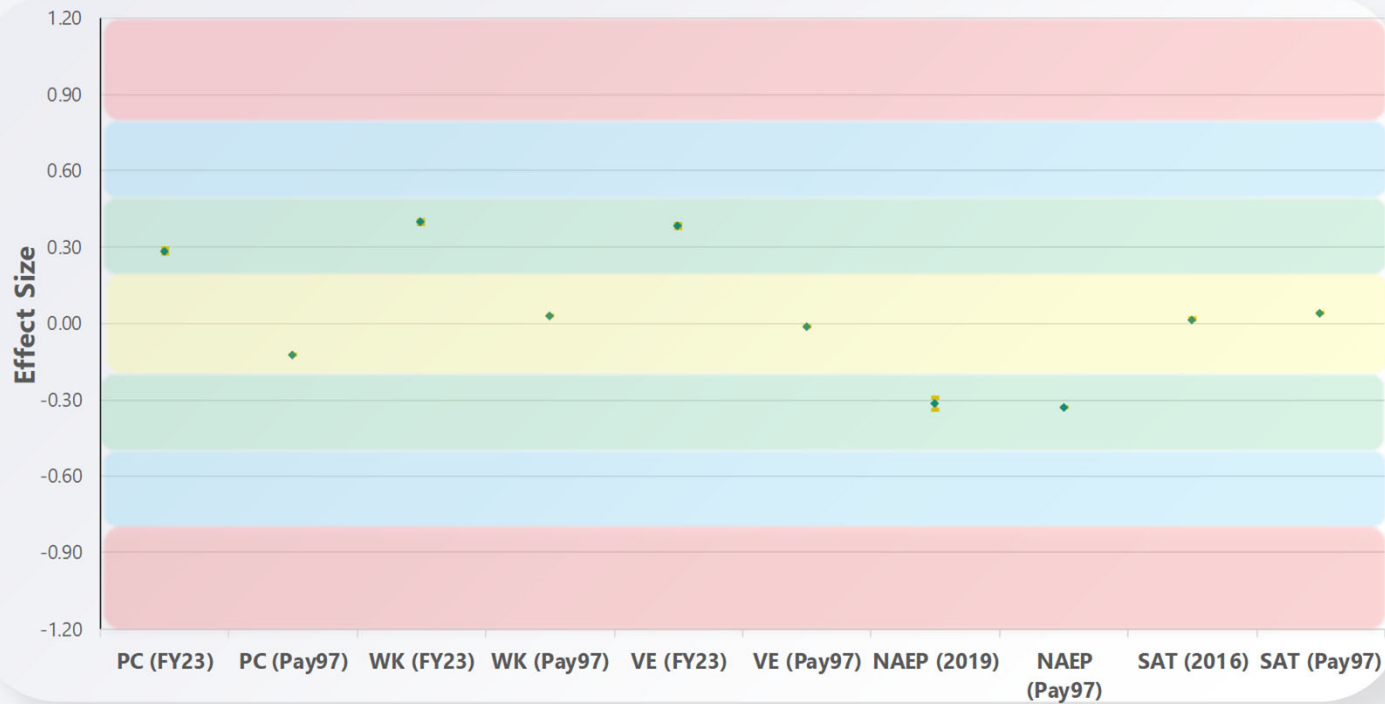
Test Effect Size

PC	0.29
WK	0.40
VE	0.38
NAEP	-0.31
SAT	0.02

Trivial Small Mod Large



## Comparison of Effect Sizes Across Testing Programs (Reading/Verbal) Female vs. Male

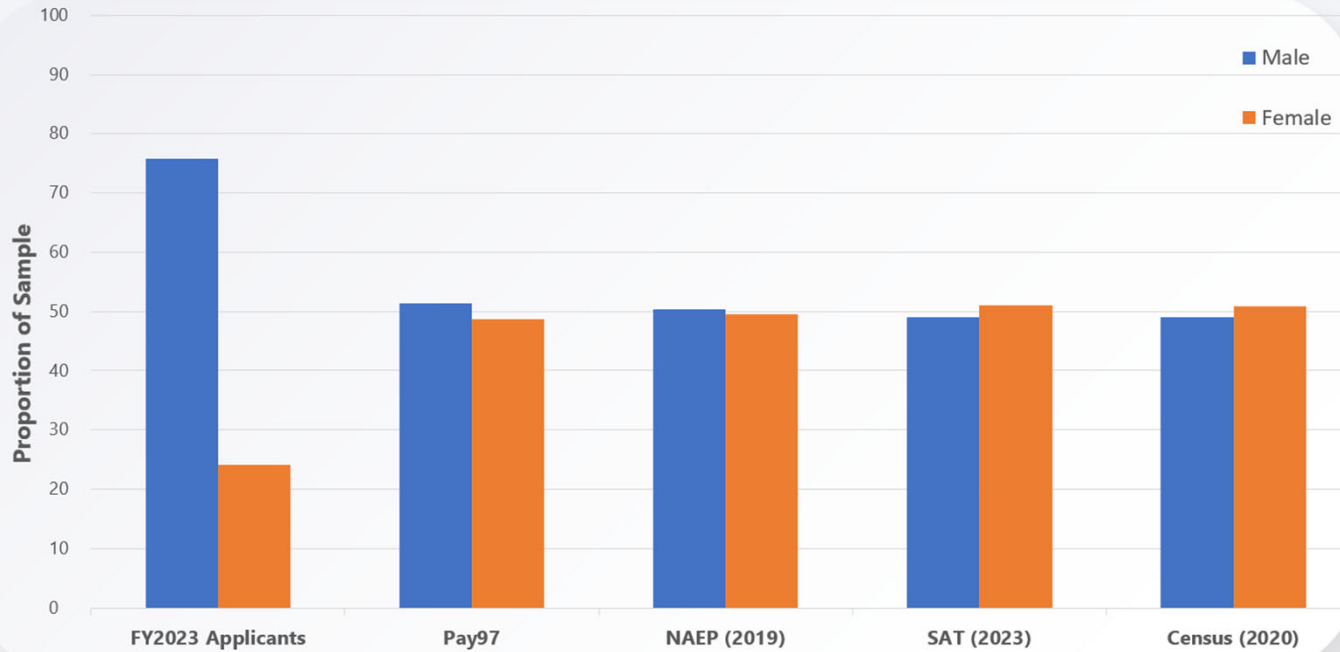


### Effect Sizes

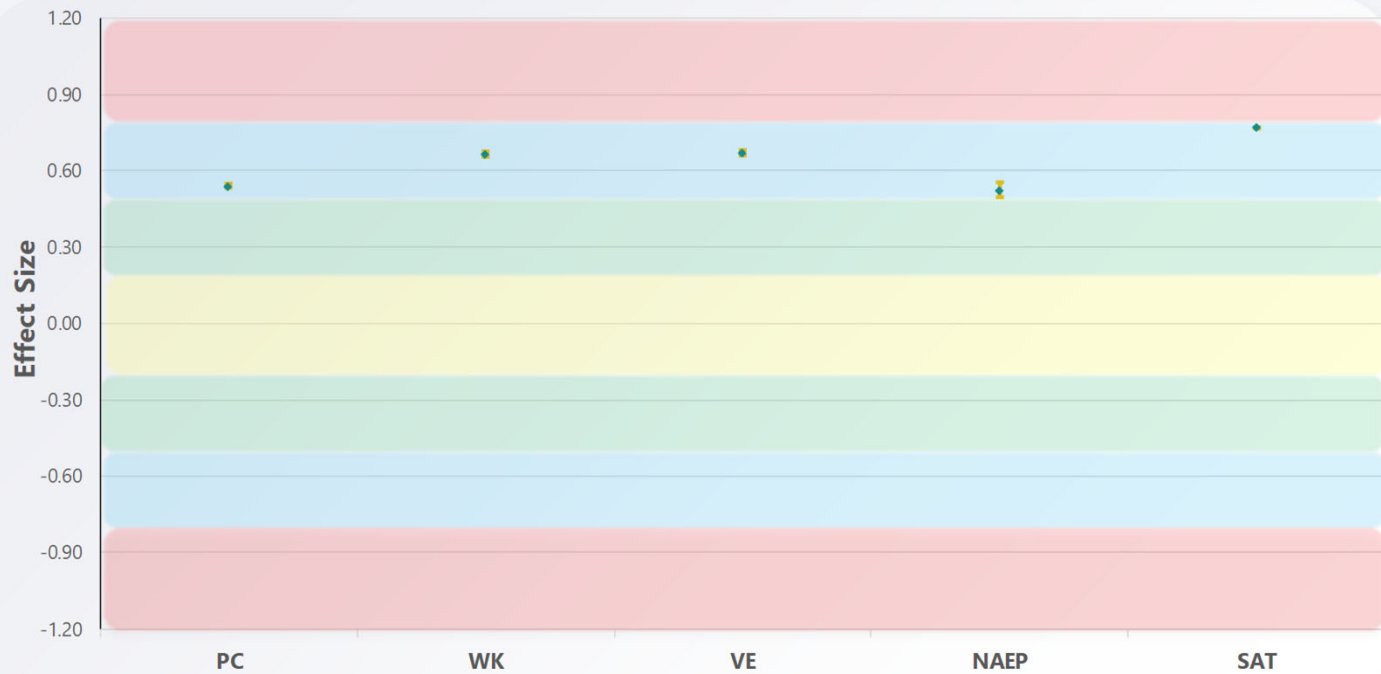
Test	Effect Size
PC (FY23)	0.29
PC (Pay97)	-0.12
WK (FY23)	0.40
WK (Pay97)	0.03
VE (FY23)	0.38
VE (Pay97)	-0.01
NAEP (2019)	-0.31
NAEP (Pay97)	-0.33
SAT (2016)	0.02
SAT (Pay97)	0.04

Trivial Small Mod Large

## Gender Representation Across Samples



# Comparison of Effect Sizes Across Testing Programs (Reading)



## Effect Sizes

Comparison

Female vs Male

**H vs NHW**

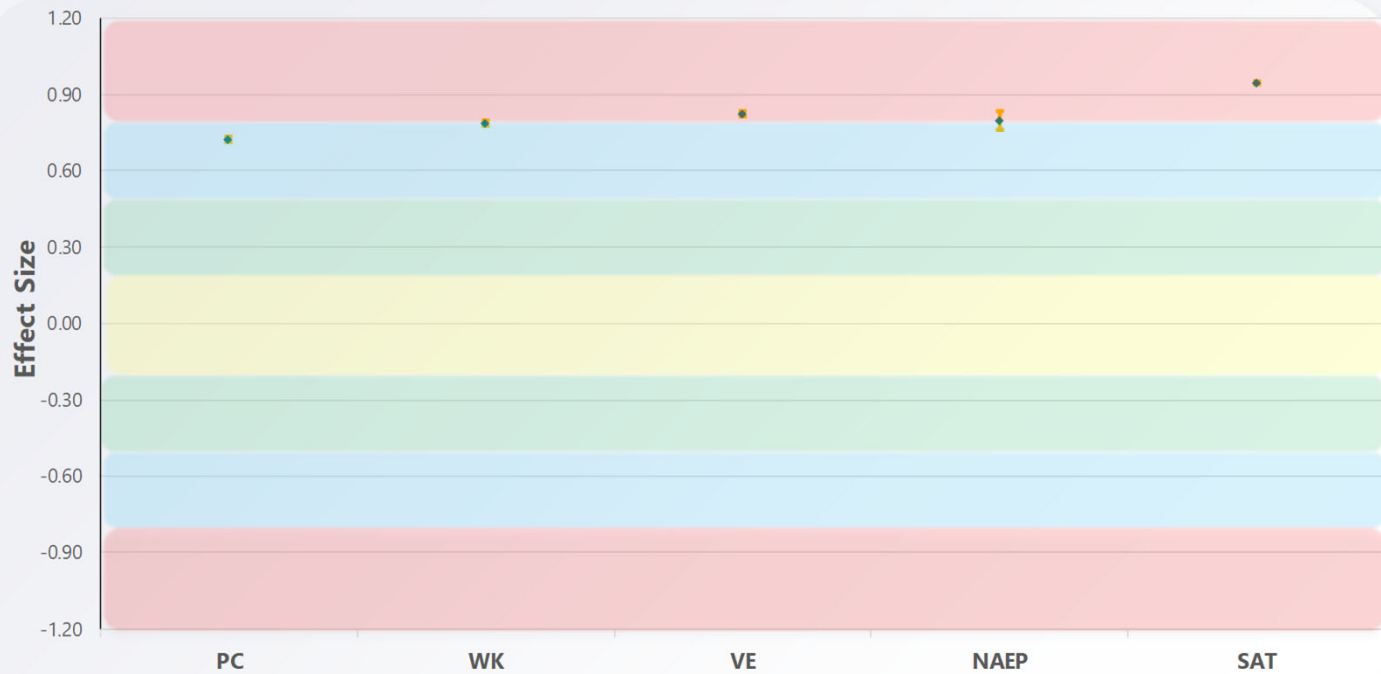
NHA vs NHW

NHB vs NHW

Test	Effect Size
PC	<b>0.54</b>
WK	<b>0.67</b>
VE	<b>0.67</b>
NAEP	<b>0.53</b>
SAT	<b>0.77</b>

Trivial Small Mod Large

# Comparison of Effect Sizes Across Testing Programs (Reading)



## Effect Sizes

Comparison

Female vs Male

H vs NHW

NHA vs NHW

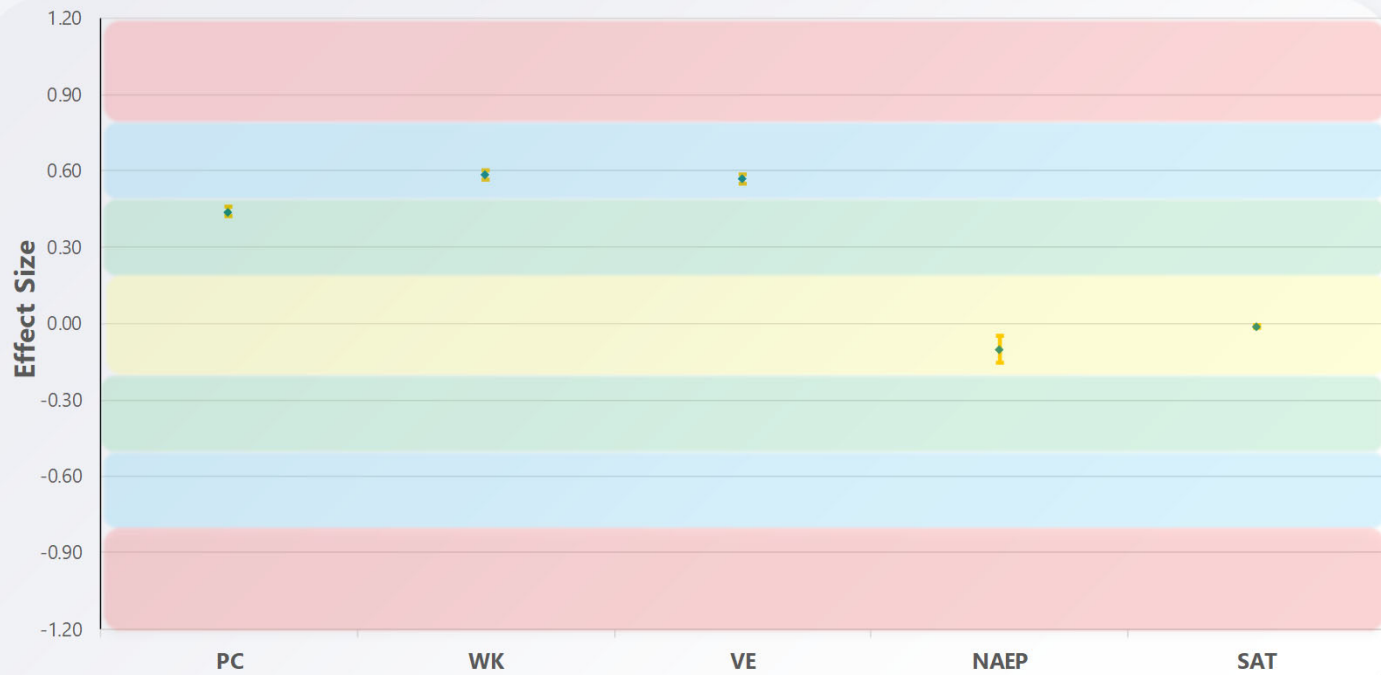
**NHB vs NHW**

Test Effect Size

PC	0.73
WK	0.79
VE	0.82
NAEP	0.80
SAT	0.95

Trivial Small Mod Large

# Comparison of Effect Sizes Across Testing Programs (Reading)



## Effect Sizes

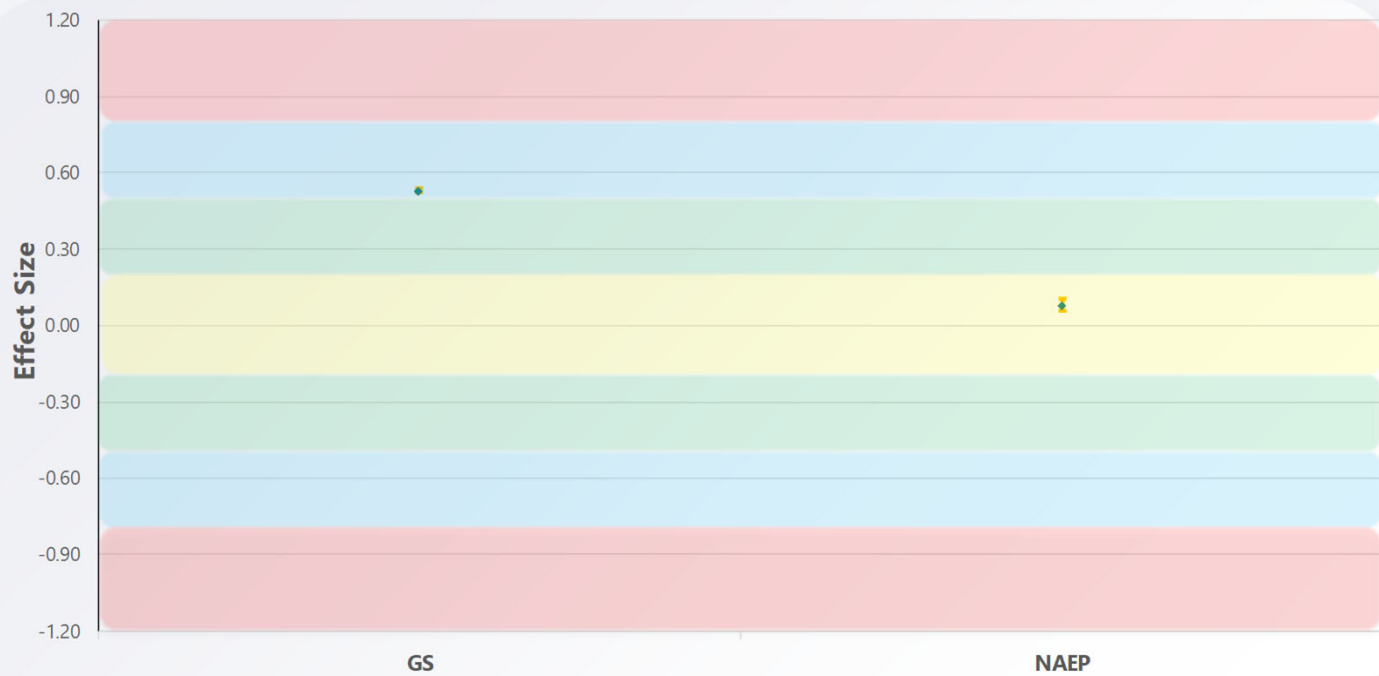
Comparison

- Female vs Male
- H vs NHW
- NHA vs NHW**
- NHB vs NHW

Test	Effect Size
PC	0.44
WK	0.58
VE	0.57
NAEP	-0.10
SAT	-0.01

Trivial Small Mod Large

## Comparison of Effect Sizes Across Testing Programs (Science)



### Effect Sizes

Comparison

Female vs Male

H vs NHW

NHA vs NHW

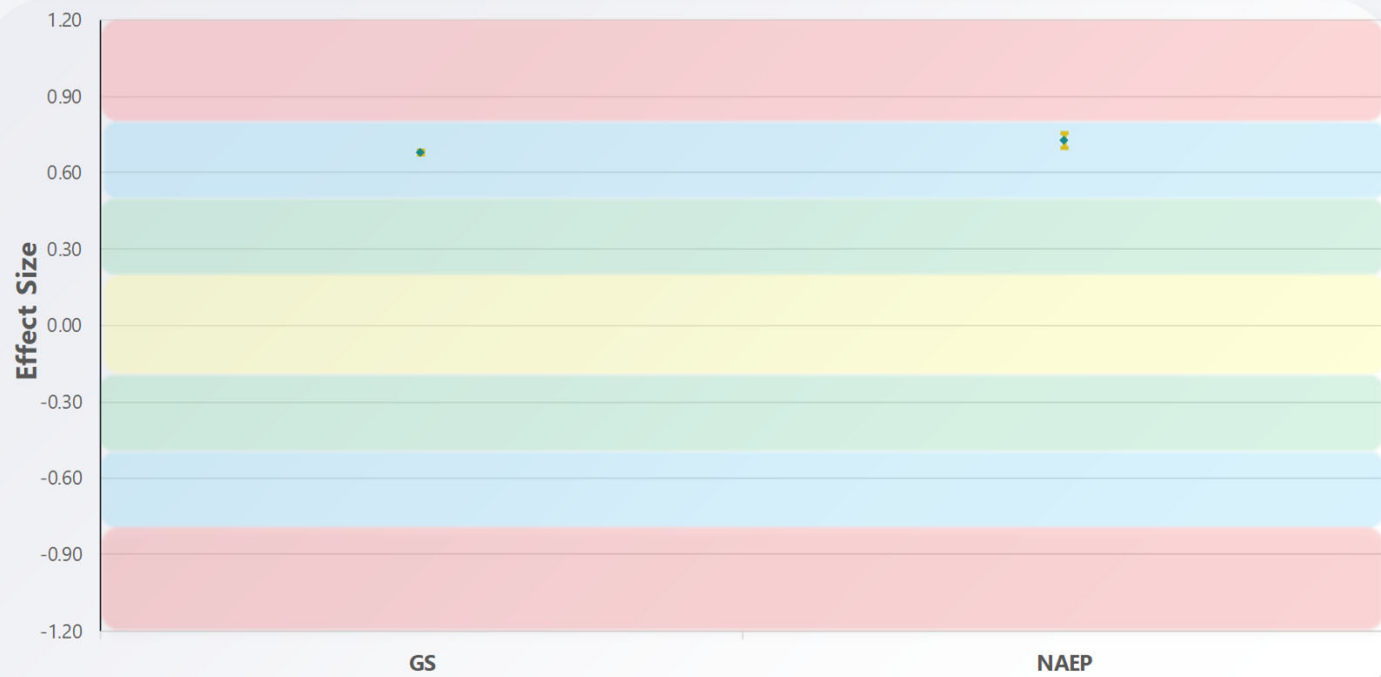
NHB vs NHW

Test Effect Size

GS	0.53
NAEP	0.08

Trivial Small Mod Large

## Comparison of Effect Sizes Across Testing Programs (Science)



### Effect Sizes

Comparison

Female vs Male

**H vs NHW**

NHA vs NHW

NHB vs NHW

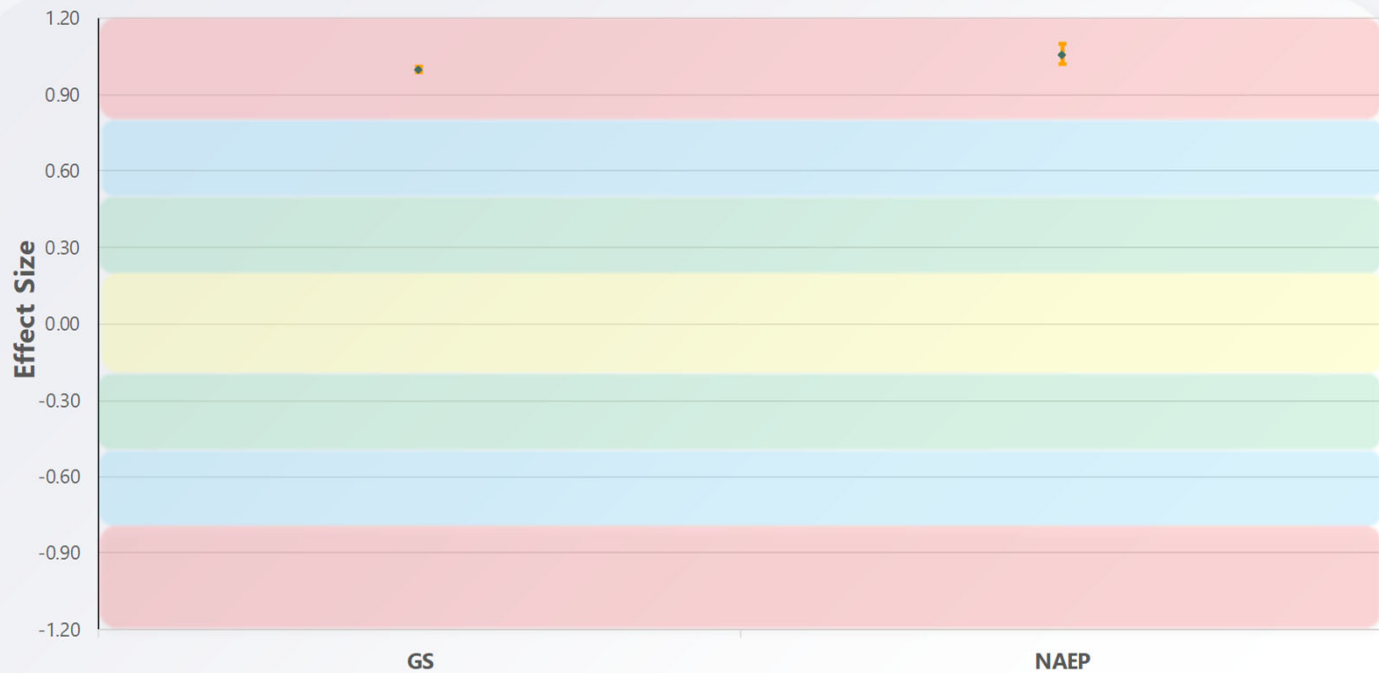
Test Effect Size

GS 0.68

NAEP 0.73

Trivial Small Mod Large

## Comparison of Effect Sizes Across Testing Programs (Science)



### Effect Sizes

Comparison

Female vs Male

H vs NHW

NHA vs NHW

**NHB vs NHW**

Test Effect Size

GS 1.00

NAEP 1.06

Trivial

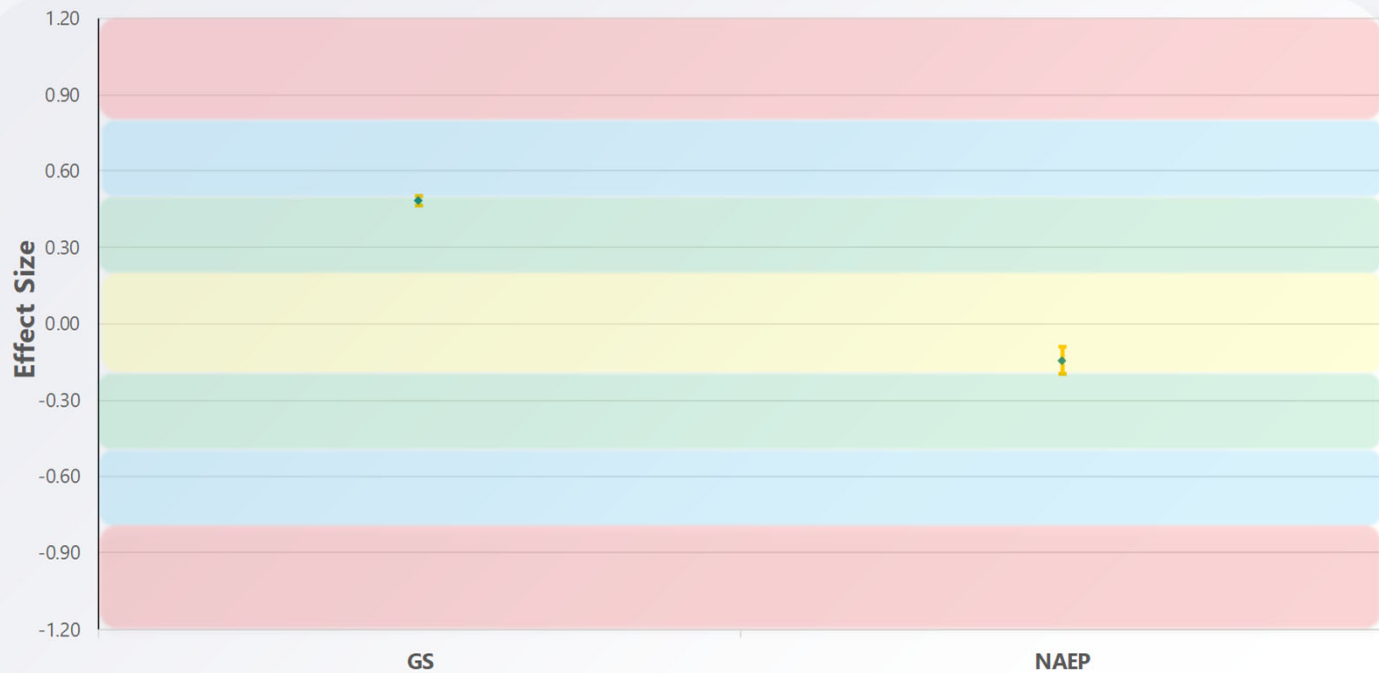
Small

Mod

Large



# Comparison of Effect Sizes Across Testing Programs (Science)



## Effect Sizes

Comparison

Female vs Male

H vs NHW

**NHA vs NHW**

NHB vs NHW

Test Effect Size

GS 0.49

NAEP -0.14

Trivial

Small

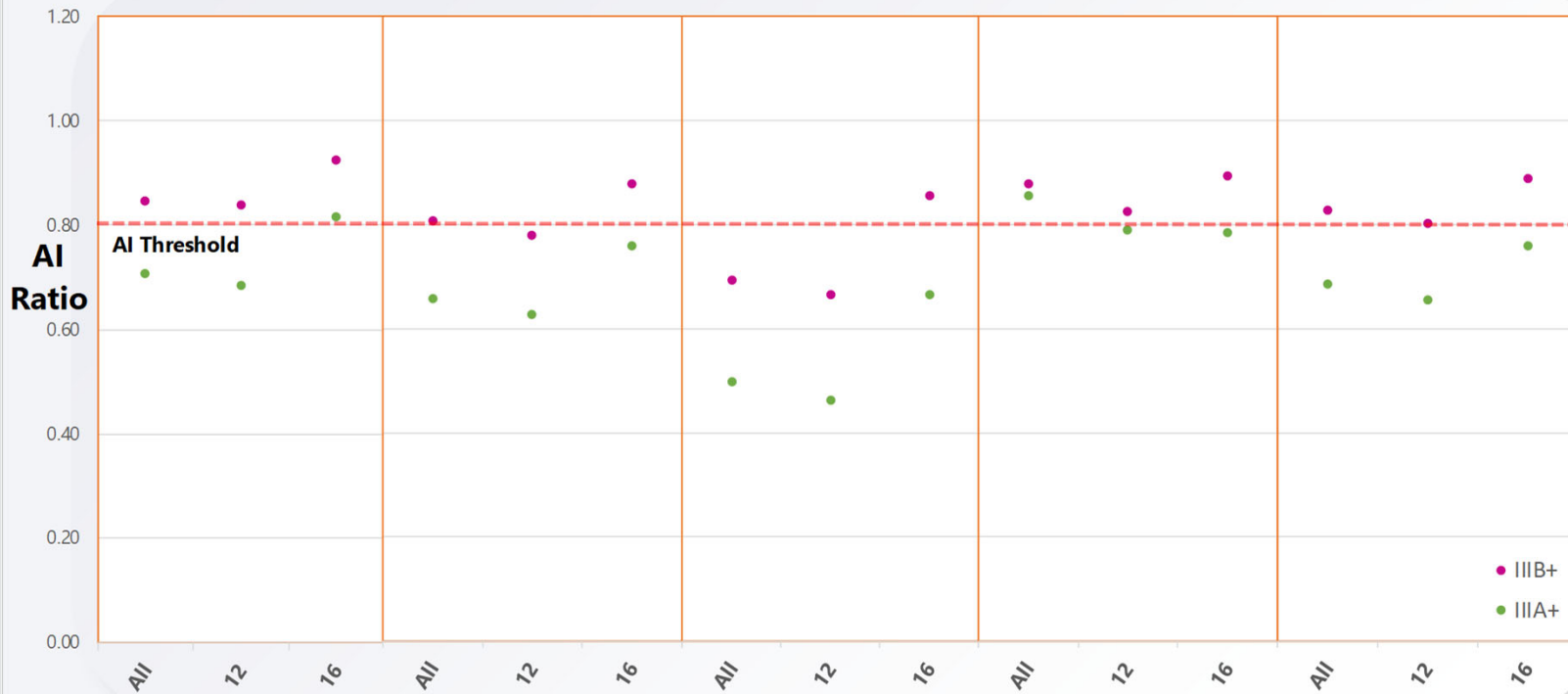
Mod

Large

---

# Impact Ratio by Education Level

## Comparison of FY2023 Impact Ratios for Years of Education Group



### Adverse Impact Ratios

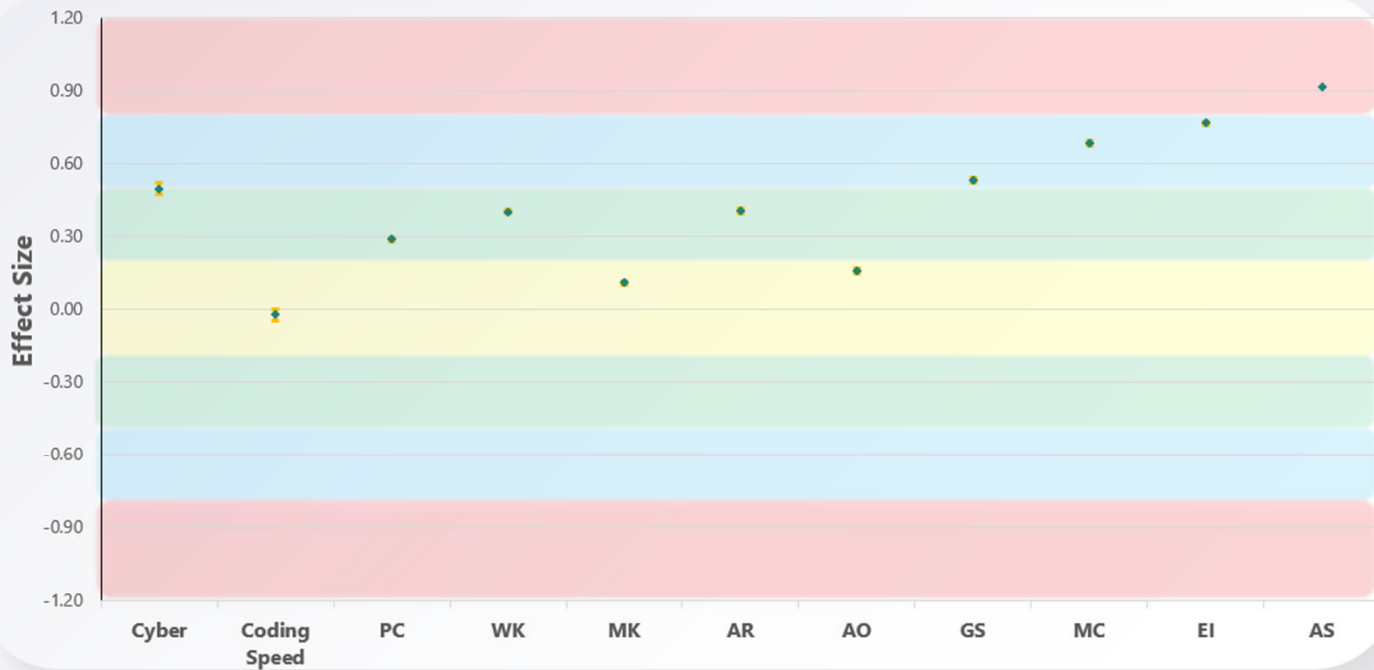
Year	Education	III B+	III A+
Female vs Male	All	0.85	0.71
Female vs Male	12	0.84	0.68
Female vs Male	16	0.93	0.82
HW vs NHW	All	0.81	0.66
HW vs NHW	12	0.78	0.63
HW vs NHW	16	0.88	0.76
NHB vs NHW	All	0.69	0.50
NHB vs NHW	12	0.67	0.46
NHB vs NHW	16	0.86	0.67
NHA vs NHW	All	0.88	0.86
NHA vs NHW	12	0.83	0.79
NHA vs NHW	16	0.89	0.78
NWH vs NHW	All	0.83	0.69
NWH vs NHW	12	0.80	0.66
NWH vs NHW	16	0.89	0.76

Trivial Small Mod Large

---

# Effect Sizes for Special Tests

# Effect Sizes for ASVAB Scores FY2023 with Cyber Test



### Effect Sizes

Comparison

**Female vs Male**

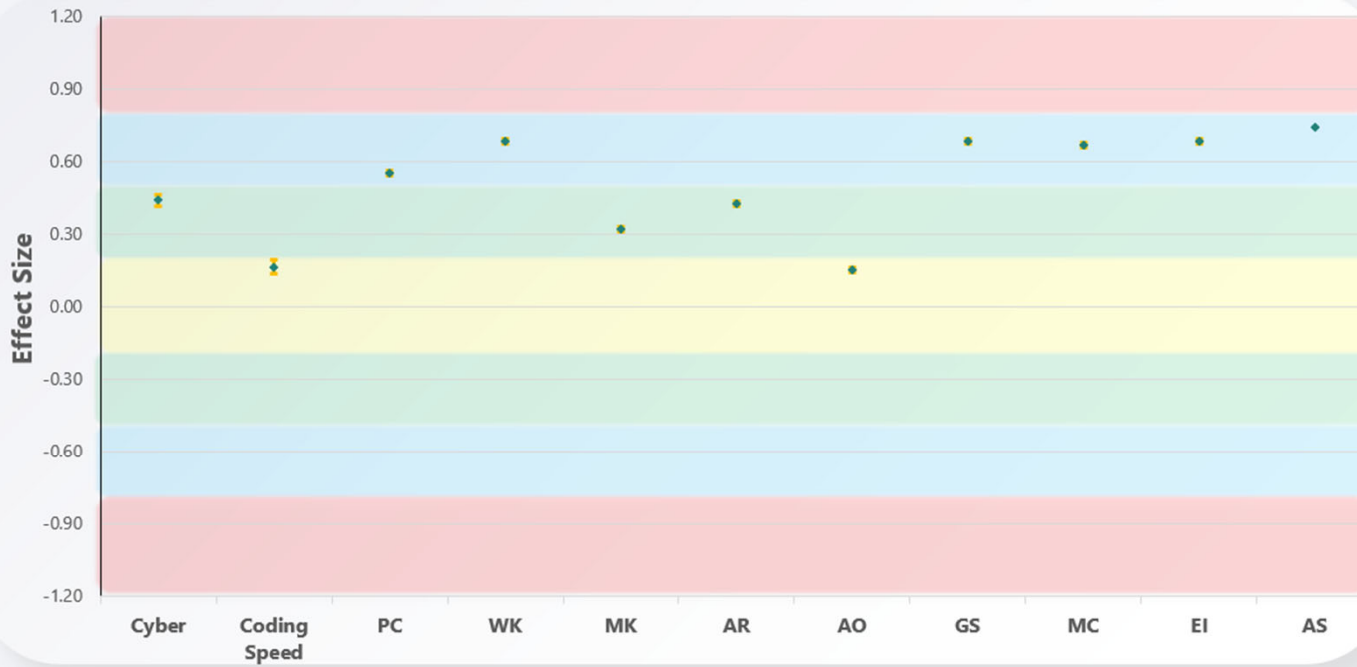
NHA vs NHW

NWH vs NHW

Test	Effect Size
Cyber	0.50
Coding Speed	-0.02
PC	0.29
WK	0.40
MK	0.11
AR	0.40
AO	0.16
GS	0.53
MC	0.68
EI	0.76
AS	0.91

Trivial
Small
Mod
Large

## Effect Sizes for ASVAB Scores FY2023 with Cyber Test



### Effect Sizes

Comparison

Female vs Male

**HW vs NHW**

NHA vs NHW

NHB vs NHW

NWH vs NHW

Test	Effect Size
Cyber	0.44
Coding Speed	0.16
PC	0.55
WK	0.68
MK	0.32
AR	0.42
AO	0.15
GS	0.68
MC	0.67
EI	0.68
AS	0.74

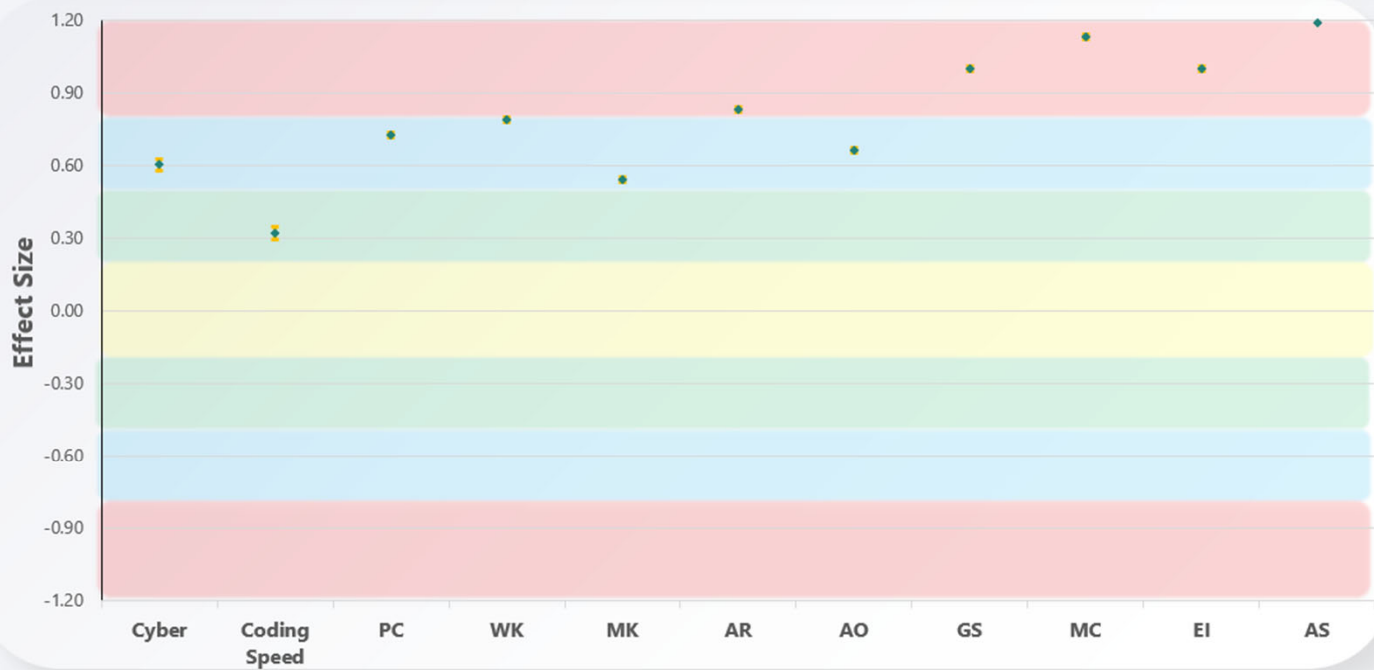
Trivial

Small

Mod

Large

## Effect Sizes for ASVAB Scores FY2023 with Cyber Test



### Effect Sizes

Comparison

Female vs Male

HW vs NHW

NHA vs NHW

**NHB vs NHW**

NWH vs NHW

Test	Effect Size
Cyber	0.60
Coding Speed	0.32
PC	0.73
WK	0.79
MK	0.54
AR	0.83
AO	0.66
GS	1.00
MC	1.13
EI	1.00
AS	1.19

Trivial Small Mod Large

## Effect Sizes for ASVAB Scores FY2023 with Cyber Test



### Effect Sizes

Comparison

Female vs Male

HW vs NHW

**NHA vs NHW**

NHB vs NHW

NWH vs NHW

Test	Effect Size
Cyber	0.15
Coding Speed	0.00
PC	0.44
WK	0.58
MK	-0.31
AR	-0.06
AO	-0.04
GS	0.49
MC	0.42
EI	0.49
AS	0.83

Trivial Small Mod Large



# Effect Sizes for ASVAB Scores FY2023 with Cyber Test



## Effect Sizes

Comparison

Female vs Male

HW vs NHW

NHA vs NHW

NHB vs NHW

**NWH vs NHW**

Test	Effect Size
Cyber	0.40
Coding Speed	0.27
PC	0.49
WK	0.59
MK	0.32
AR	0.49
AO	0.23
GS	0.68
MC	0.73
EI	0.70
AS	0.84

Trivial Small Mod Large

---

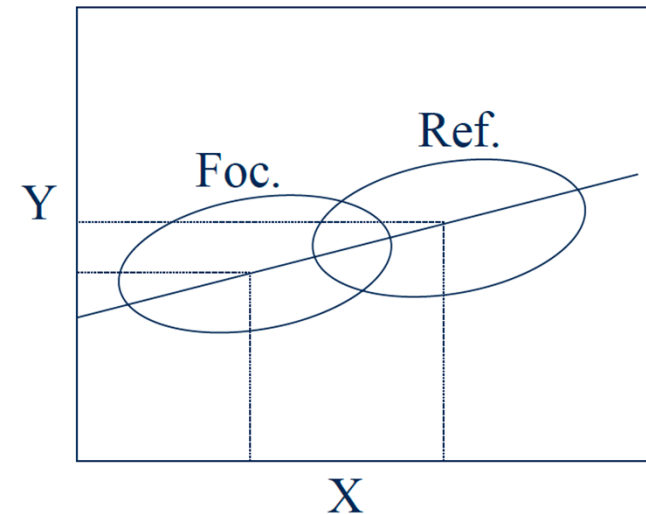
# Conclusions

## Conclusions and Caveats

- For the AFQT tests and GS, the direction and magnitude of overall impact is generally consistent with comparable SAT and NAEP tests, which suggests that impact on ASVAB tests is reflective of differences in job or training performance
  - Comparisons across programs may be somewhat restricted due to differences in group definitions, testing populations, test content, etc.
    - NAEP is effectively an unrestricted sample
    - Those self-selecting into the Armed Services likely differ from SAT test-takers in terms of personality, motivation, and other characteristics

## Conclusions and Caveats

- Adverse impact does **not** reflect test bias if validity research shows that the test is equally valid for relevant groups.
  - Historically, a regression-based approach has been advocated to evaluate the existence of bias. Lack of test bias is indicated when the regression line relating the test score [X] and a criterion [Y] is the same for each group.
  - This was the approach taken by Putka et al. (2022).



From Ghiselli, Campbell, & Zedeck. (1981). *Measurement Theory for the Behavioral Sciences*.

## Conclusions for Special Tests

- Cyber Test and Coding Speed generally exhibited small-to-moderate effects and were usually as low or lower than most ASVAB tests
  - Effects for CT and CS were also generally consistent with those found in FY21
    - Exception: CS NHW-NHB ES in FY21 was near 0, but was near .30 in FY23
- CS usually had very small effects (ranging from 0 to 0.30)

## Questions for the DAC

- Does the DAC have any general feedback or recommendations based on these results?
- For future analyses, are there any other results the DAC would be interested in seeing?

## Acknowledgments

- Co-PD: Jessica Johnston-Fisher
- Analyst Team: Vanessa Nguyen, Eryn Nielsen
- Other HumRRO Contributors: Matt Reeder, Insu Paek, Emily Borawski, Cathedia Rose, Jeff Dahlke
- DTAC: Matt Trippe, Greg Manley, Ping Yin, Mary Pommerich, Liz Waterbury, Tom Waterbury



---

# Thank You!

For more information  
please contact:

Nicholas Howald  
nhowald@humrro.org  
703.881.6044