OFFICE OF PEOPLE ANALYTICS

(Adverse) Impact of the ASVAB and Special Tests: Findings for Fiscal Year 2019 Applicants

Defense Personnel Assessment Center

Ping Yin
Greg Manley
Mary Pommerich

DACMPT Meeting
September 18, 2020

POTENTIAL FOR 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 per se. However, AI may occur when a test's scores are used as the bases for selection.
- A selection test may contribute potential for 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 gives us an index to examine a test's potential for AI.

OPA
OFFICE OF PEOPLE ANALYTICS

HOW IS ADVERSE IMPACT ASSESSED?

- The four-fifths rule is often used to determine the occurrence of adverse impact:

  "A selection rate for any race, sex, or ethnic group which is less than four-fifths (or eighty percent) 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*:

$$IR = \frac{SR_{Foc}}{SR_{Ref}},$$ where SR is the selection rate

OPA
OFFICE OF PEOPLE ANALYTICS

HOW IS ADVERSE IMPACT ASSESSED?

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

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

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

OPA
OFFICE OF PEOPLE ANALYTICS

HOW IS ADVERSE IMPACT ASSESSED?

- The four-fifths rule and accompanying statistics are applied to the ASVAB by comparing qualification rates across the focal and reference groups of interest with regard to
  - 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); and
  - adverse impact, assessed using initial test scores only (i.e., scores from retests or confirmation tests are excluded from the analyses).
- Note that significance testing is not necessarily useful for analyses with very large numbers of applicants (i.e., >2000).

OPA
OFFICE OF PEOPLE ANALYTICS

POTENTIAL FOR ADVERSE IMPACT

- Effect sizes (i.e., standardized mean differences) provide a method of evaluating potential for adverse impact across individual ASVAB and Special Tests, where no direct selection occurs.
- Effect sizes are computed for all group comparisons as:

$$ES = \frac{\mu_R - \mu_F}{\sigma_p}$$

where:

μ_R is the mean score in the Reference (Majority) group.

μ_F is the mean score in the Focal (Minority) group.

σ_p is the pooled standard deviation across the two groups.

Note: Positive values are the direction of minority impact.

OPA
OFFICE OF PEOPLE ANALYTICS

CONFIDENCE INTERVALS ABOUT EFFECT SIZES

- A 95% confidence interval (δ_L, δ_U) for the effect size (ES) is computed as (Hedges & Olkin, 1985):

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

where:

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

- Effect sizes can be plotted and classified with respect to Cohen's (1988) standards of evaluation.
  - Small effect sizes start at 0.20.
  - Moderate effect sizes start at 0.50.
  - Large effect sizes start at 0.80.

OPA
OFFICE OF PEOPLE ANALYTICS

WHO IS AFFECTED BY ADVERSE IMPACT?

- The ASVAB testing program evaluates (adverse) impact for the following pairs of groups:

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

- The focal group is potentially disadvantaged relative to the reference group.
- Pairs 1–3 are the same groups that are used in evaluating DIF. Pair 4 is also included because Non-Hispanic Asians now represent >2% of the applicant population.

OPA
OFFICE OF PEOPLE ANALYTICS

WHEN IS ADVERSE IMPACT MEASURED?

- Ideally, adverse impact is assessed on a regular basis.
- Here, adverse impact is measured for applicants testing in fiscal year 2019 (FY2019 = Oct 1, 2018–Sept 30, 2019)
- Previously, adverse impact was evaluated for applicants testing in

  FY2017 = October 1, 2016 – September 30, 2017

  FY2015 = October 1, 2014 – September 30, 2015

  FY2013 = October 1, 2012 – September 30, 2013

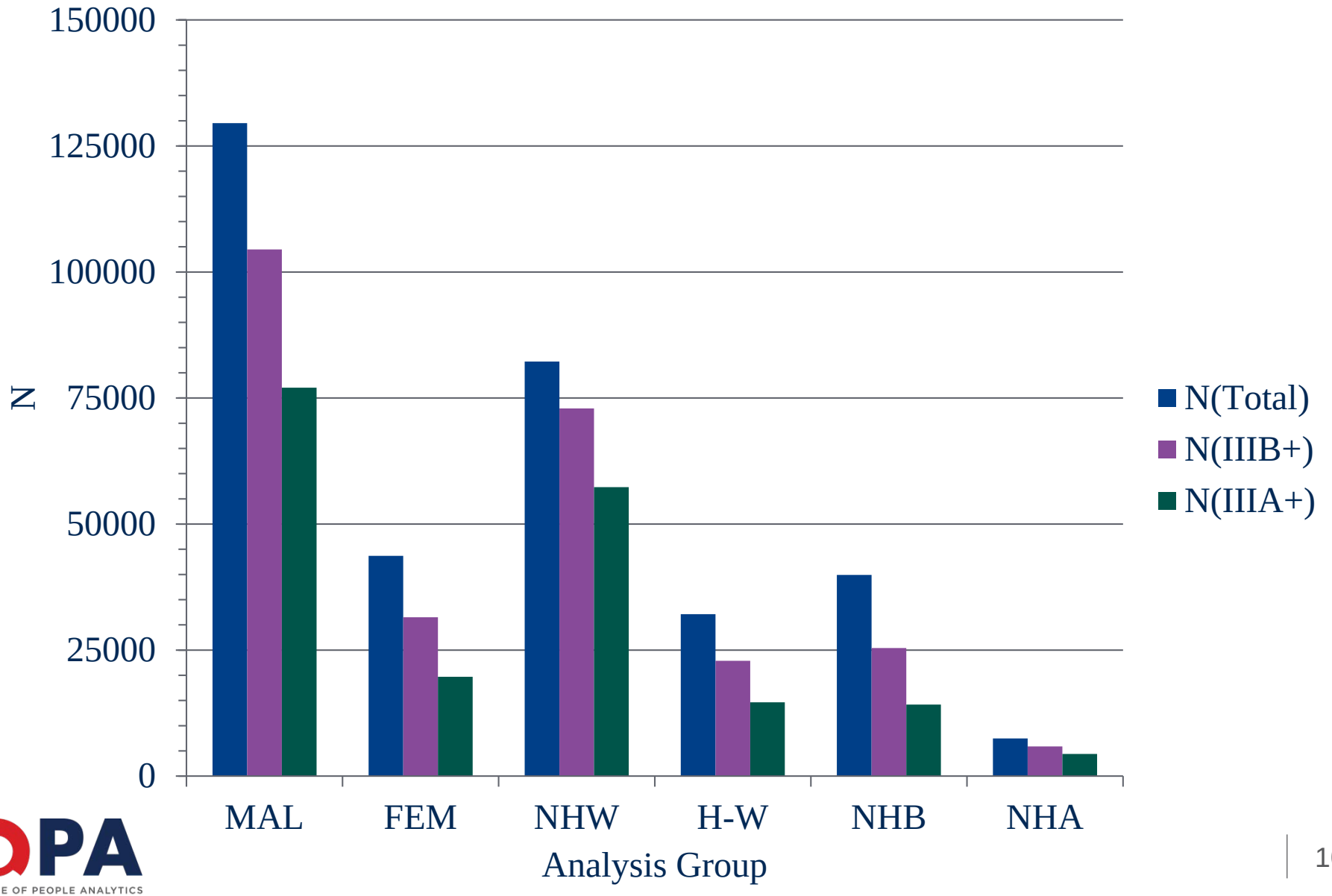
  FY2011 = October 1, 2010 – September 30, 2011

  FY2009 = October 1, 2008 – September 30, 2009

  FY2005 = October 1, 2004 – September 30, 2005

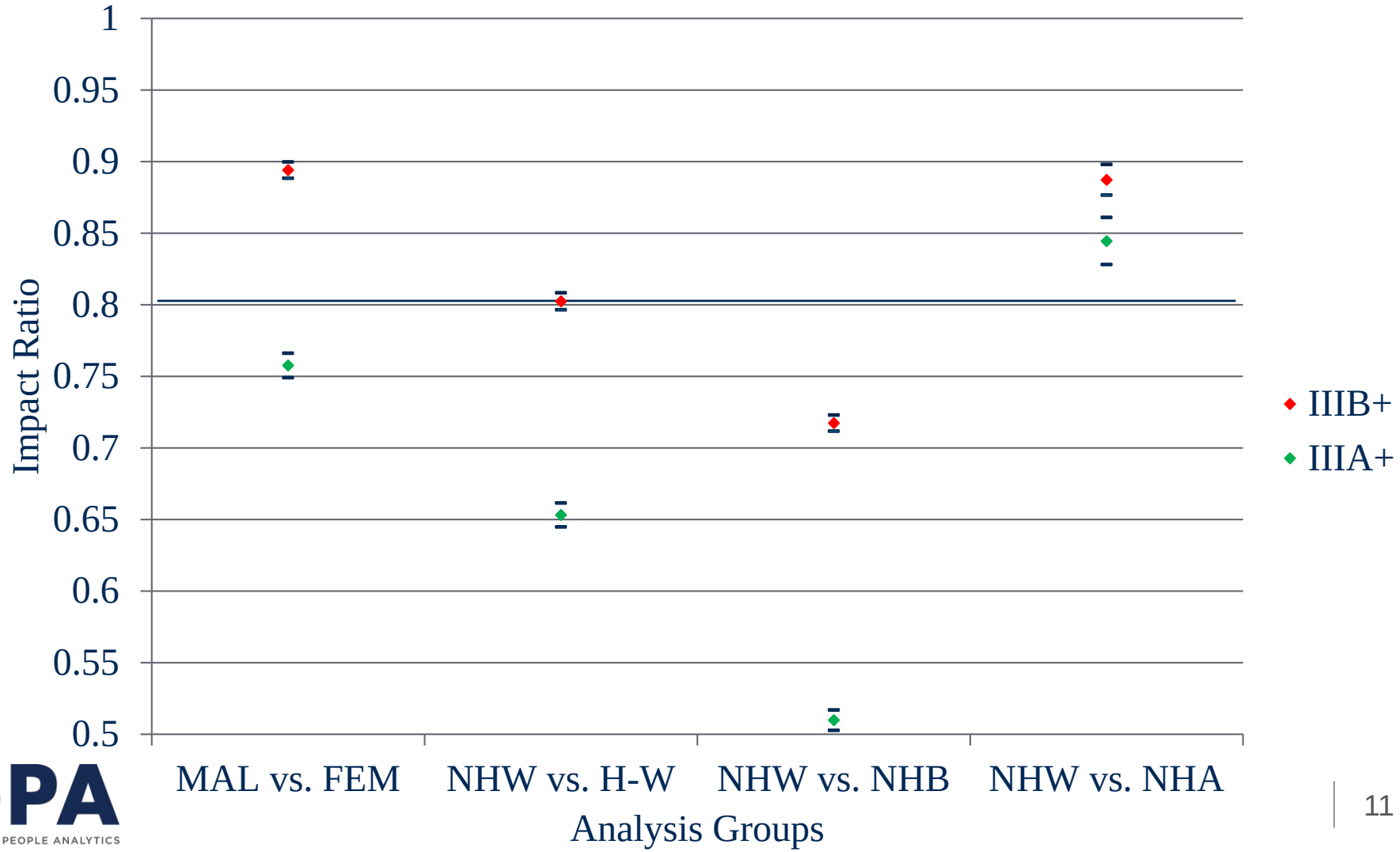
OPA
OFFICE OF PEOPLE ANALYTICS

Adverse Impact Analysis Sample Sizes FY2019



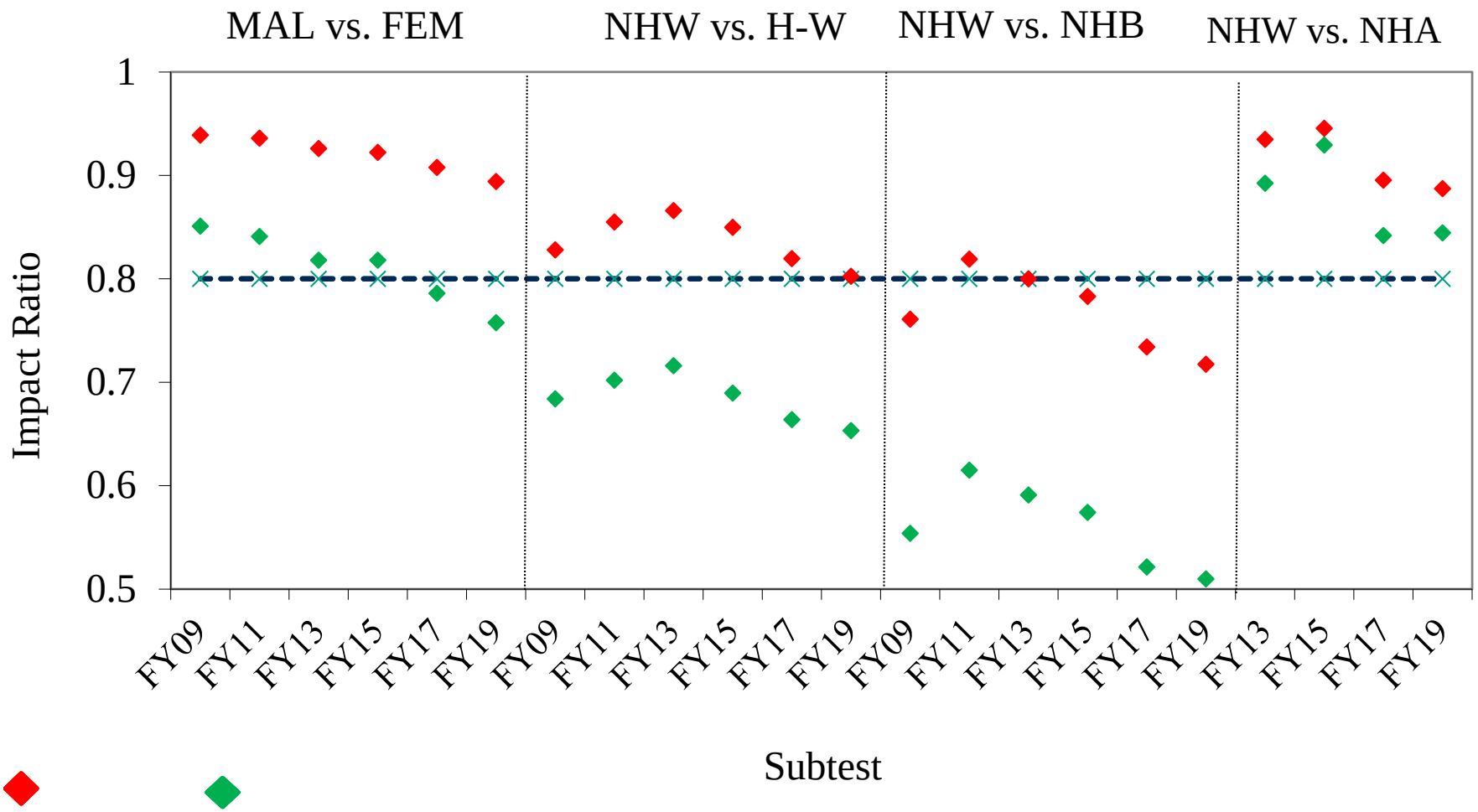
OPA
OFFICE OF PEOPLE ANALYTICS

Impact Ratio (and 95% Confidence Interval) for AFQT Cut Scores FY2019 IIIB+ & 111A+ (all education levels)



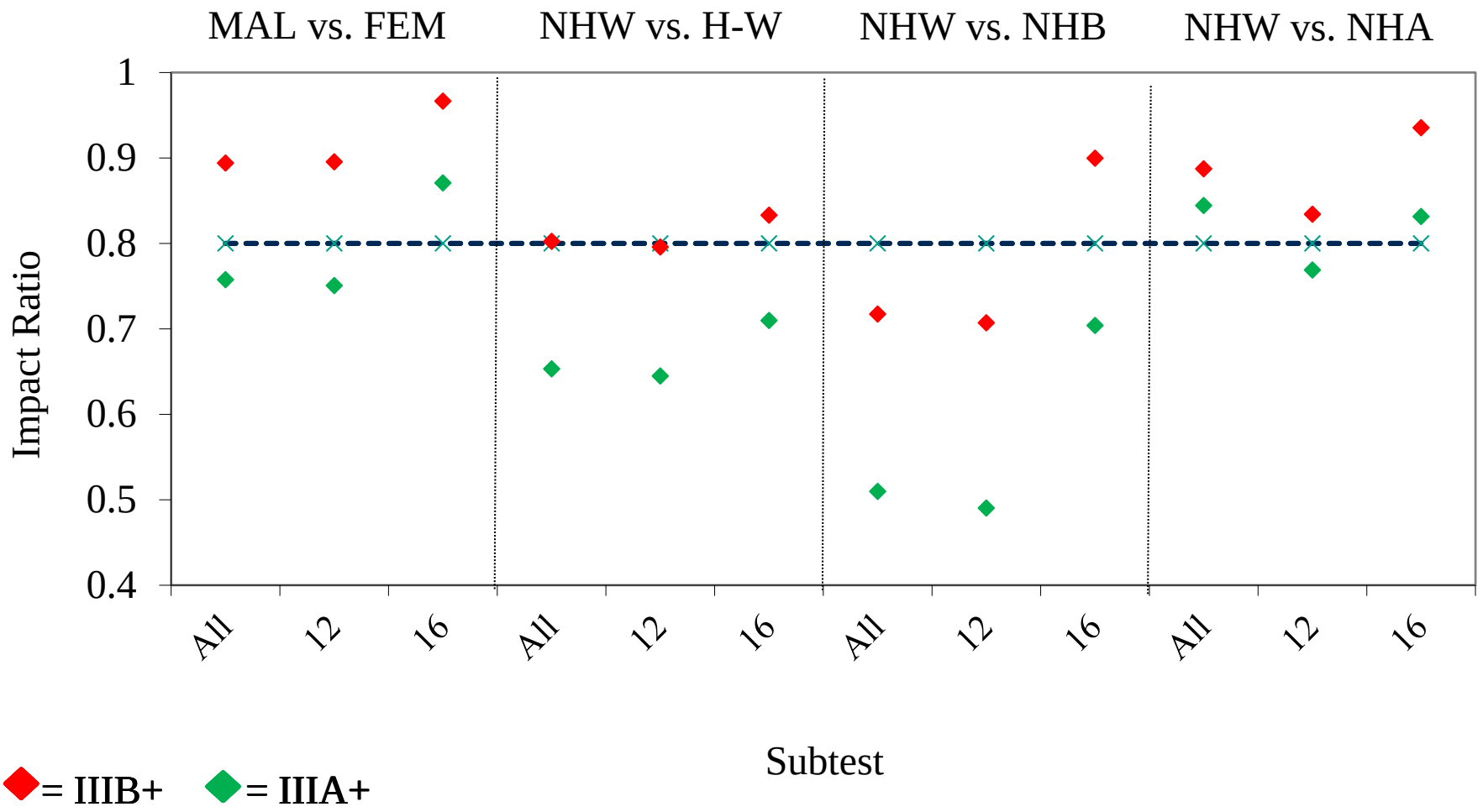
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Impact Ratios for FY09, FY11, FY13, FY15, FY17, FY19



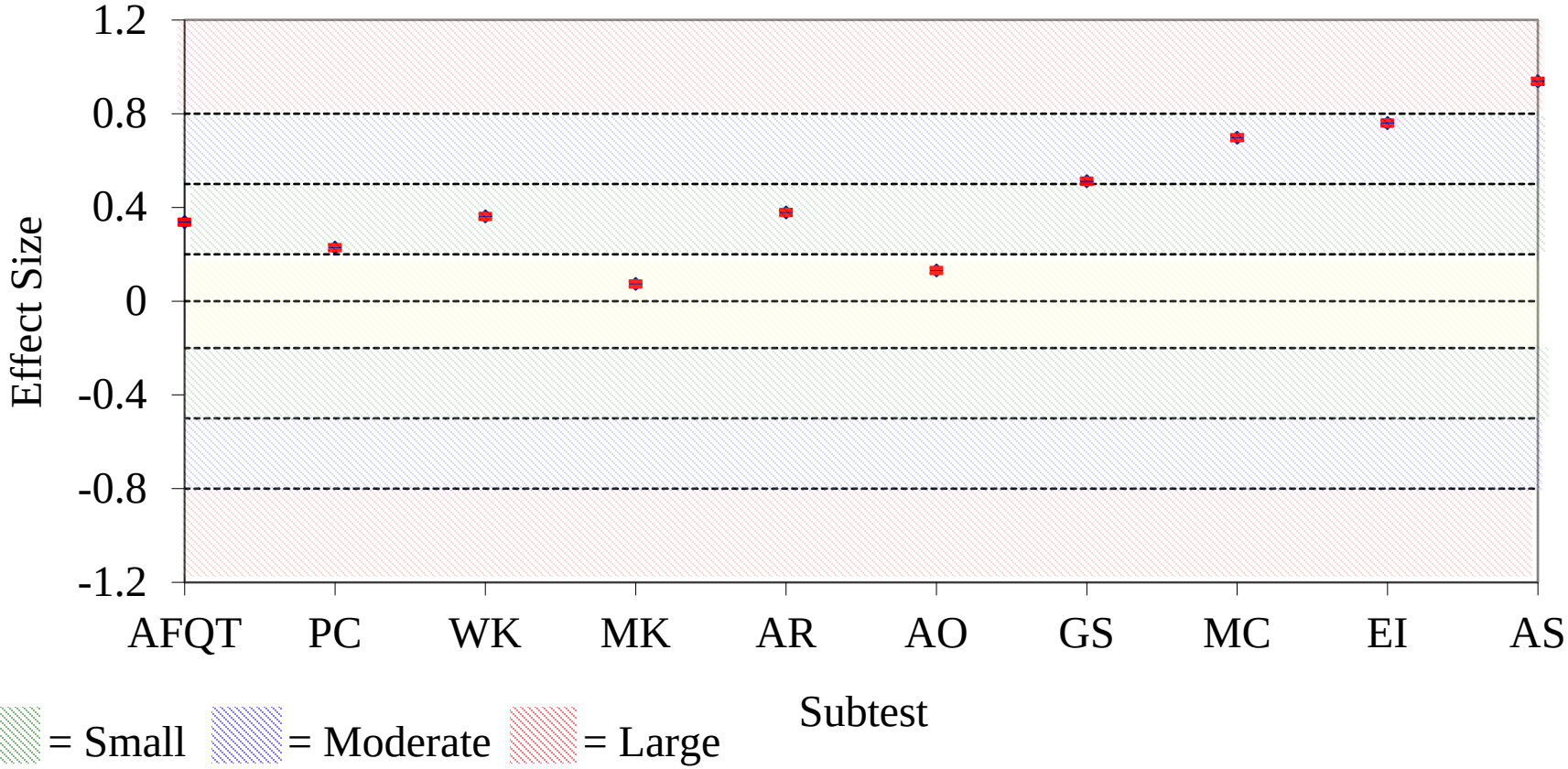
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of FY2019 Impact Ratios for Years of Education Group



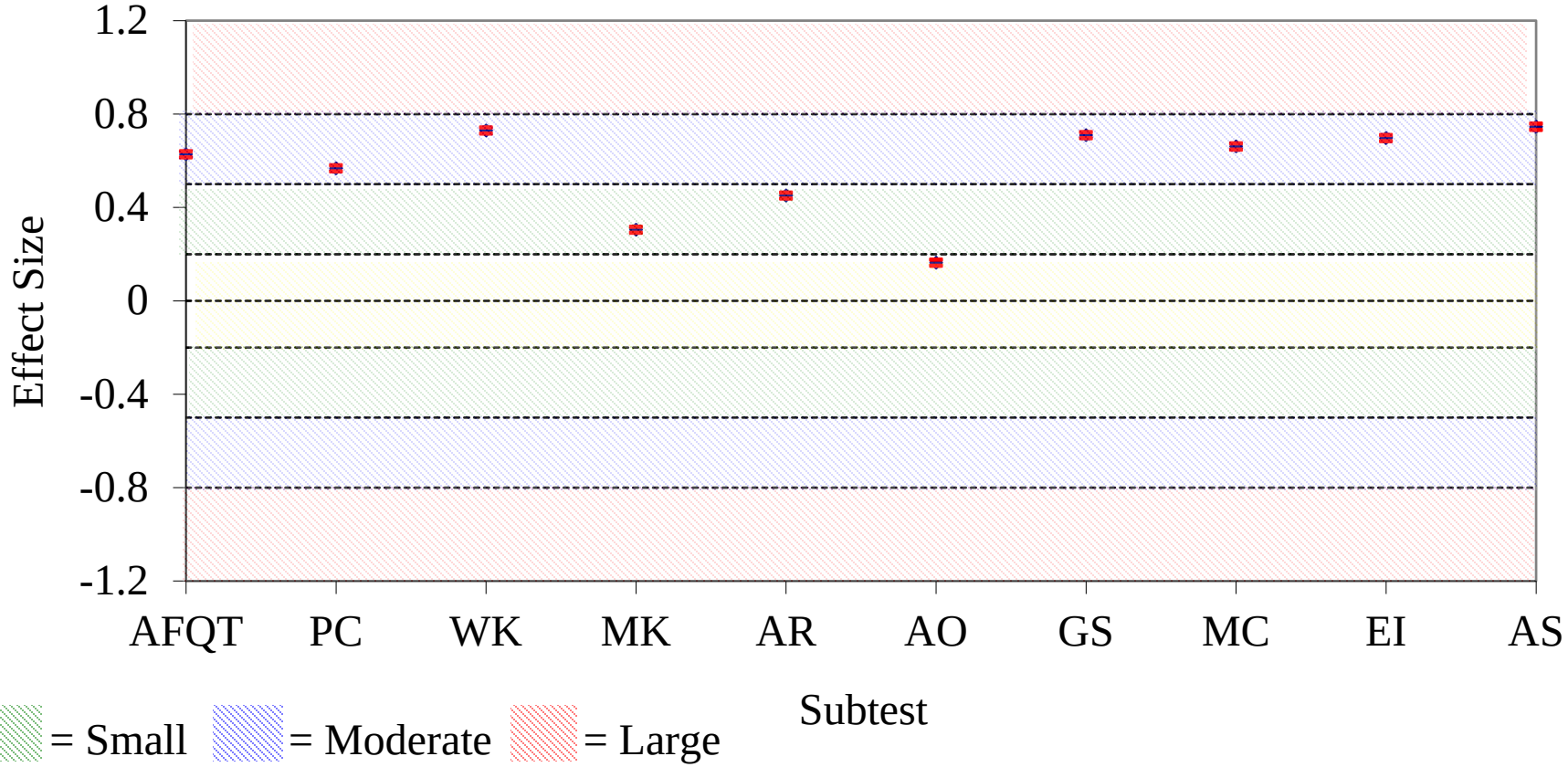
12 = 12 years of education reported; 16 = 16 years of education reported.

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Males Versus Females
FY2019



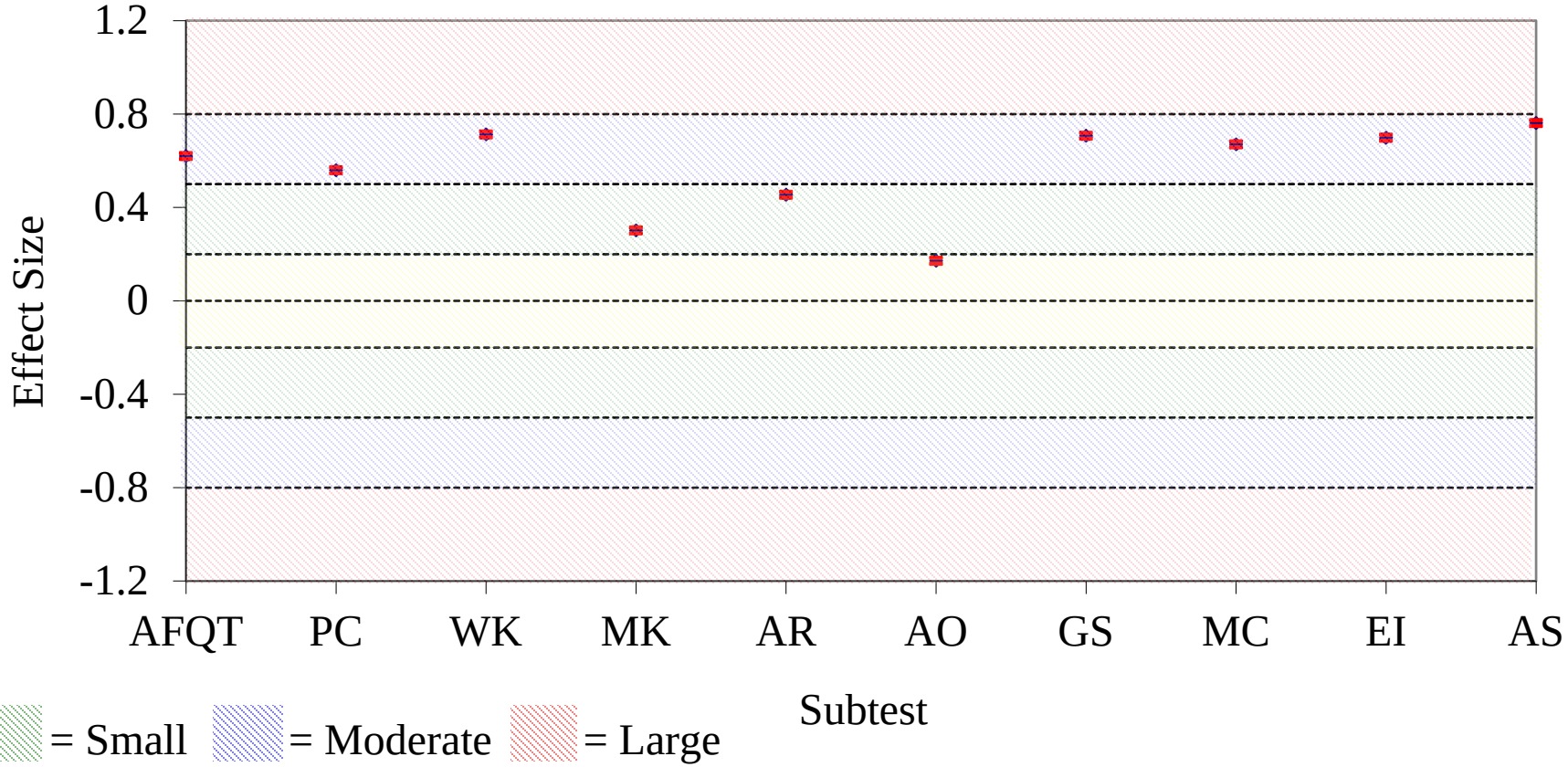
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Hispanic Whites
FY2019



OPA
OFFICE OF PEOPLE ANALYTICS

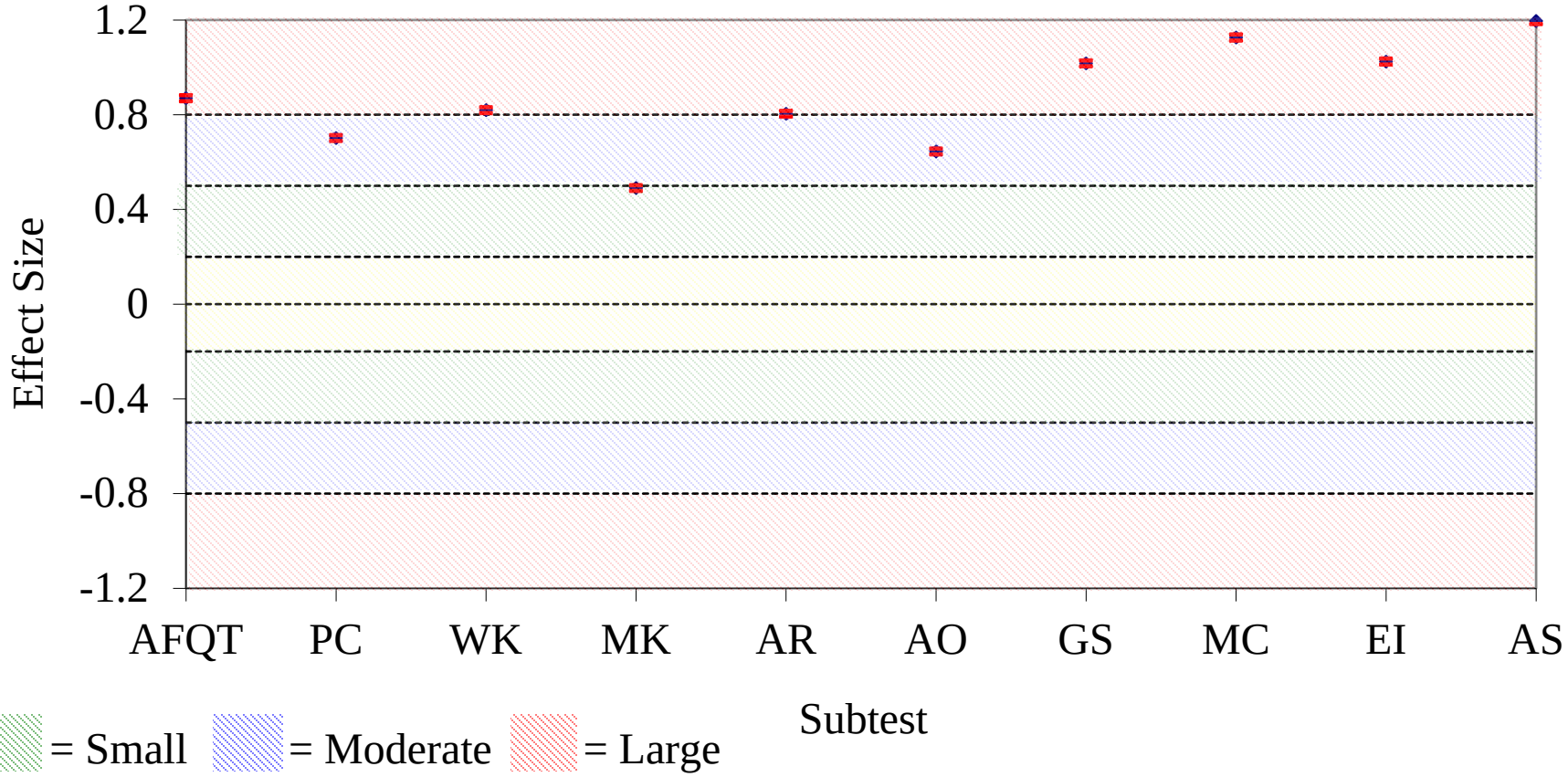
Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Hispanics*
FY2019



*All applicants who check the Hispanic box, regardless of race. Included for later comparisons with NAEP and SAT using Hispanics.

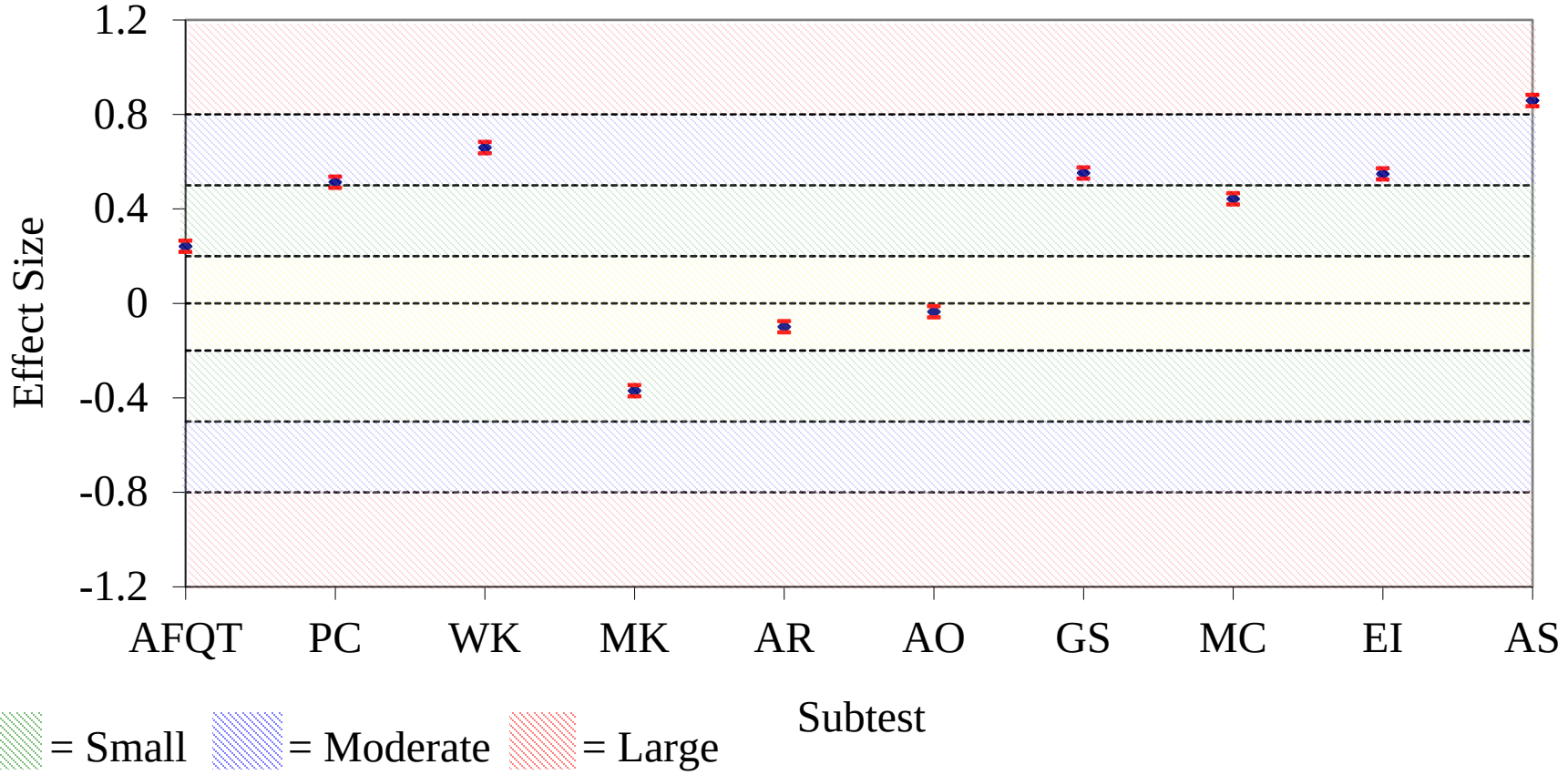
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Non-Hispanic Blacks
FY2019



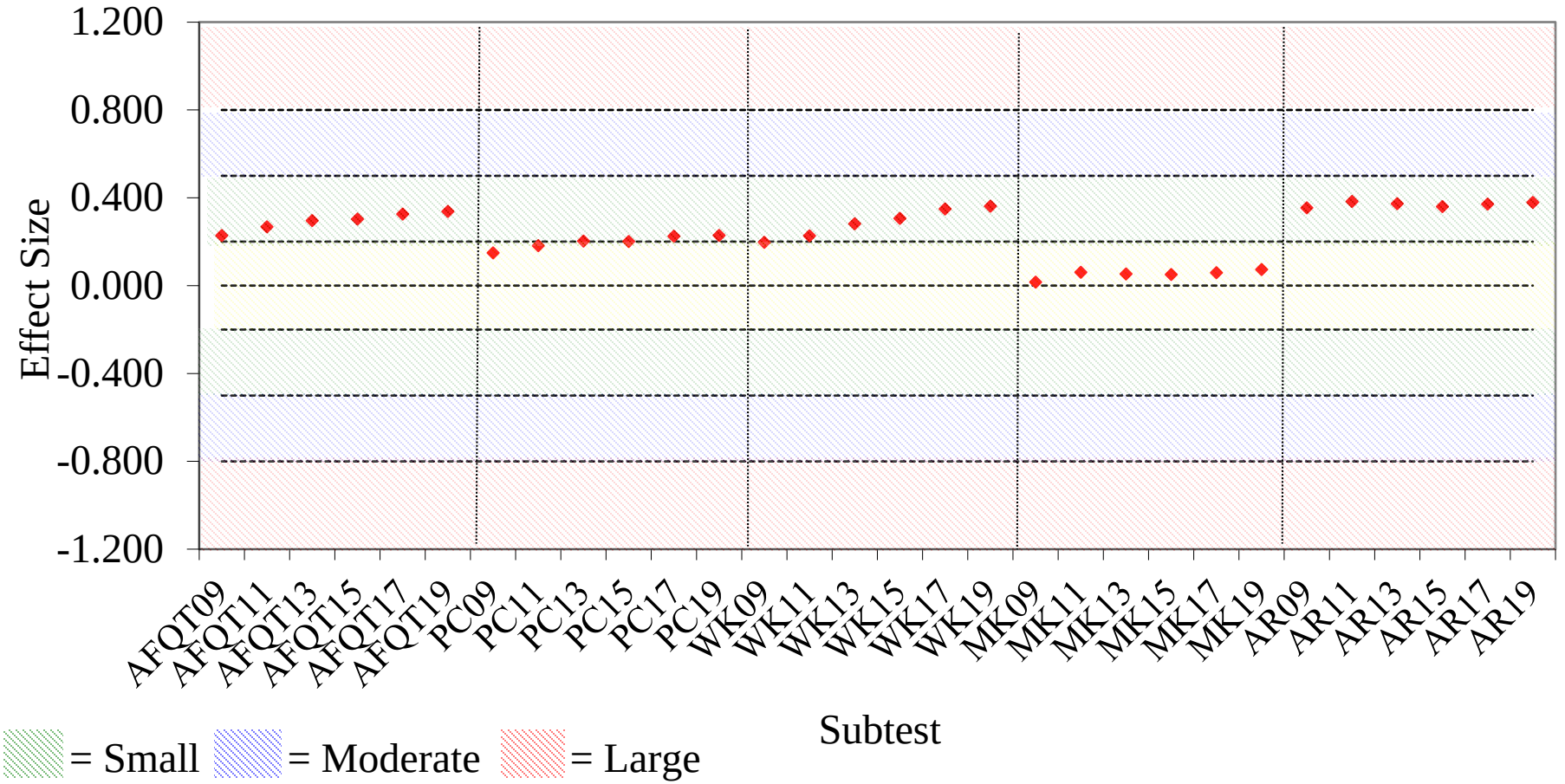
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Non-Hispanic Asians
FY2019



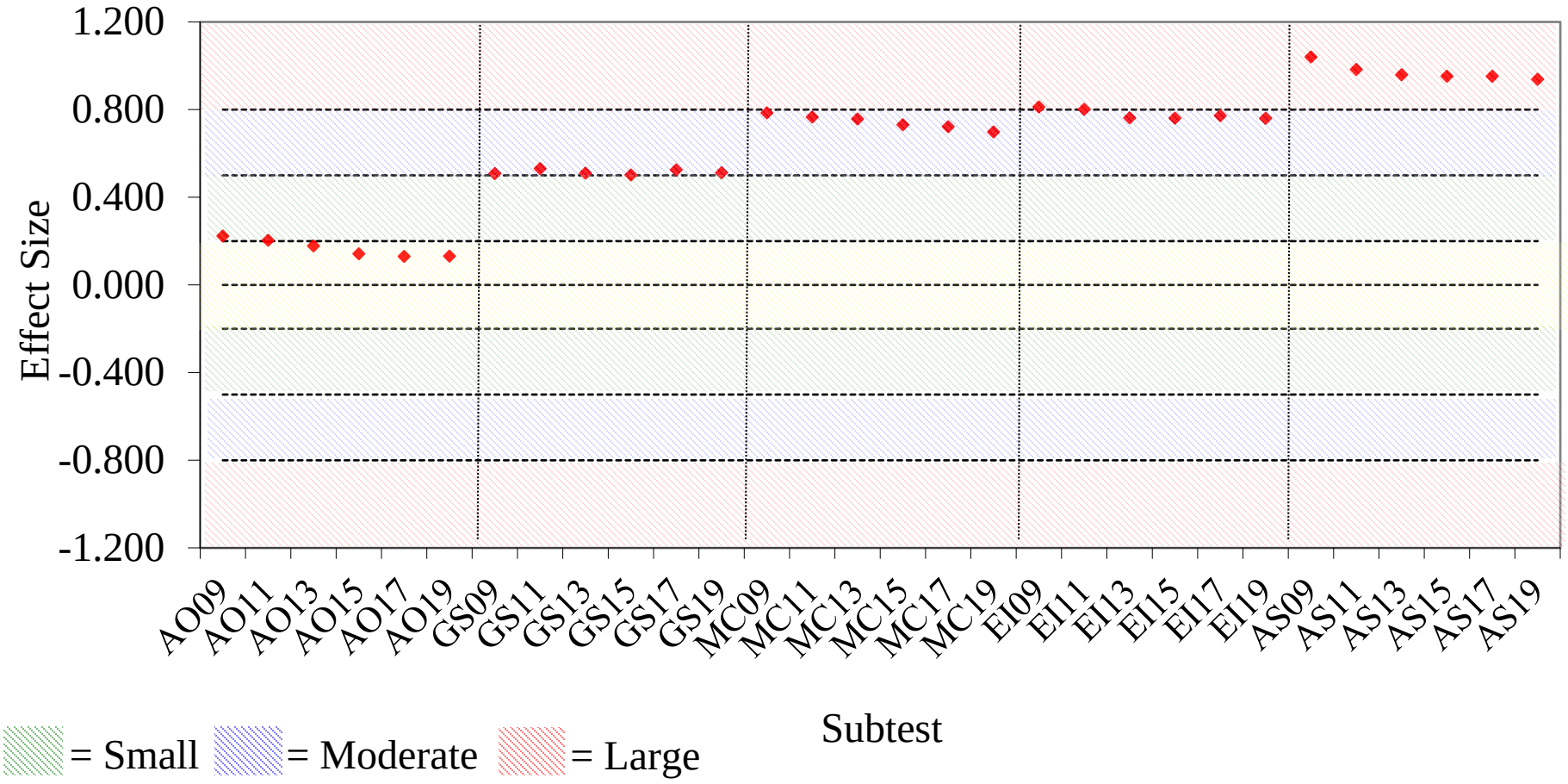
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes for FY09, FY11, FY13, FY15, FY17, FY19
Males Versus Females
AFQT Tests/Score



OPA
OFFICE OF PEOPLE ANALYTICS

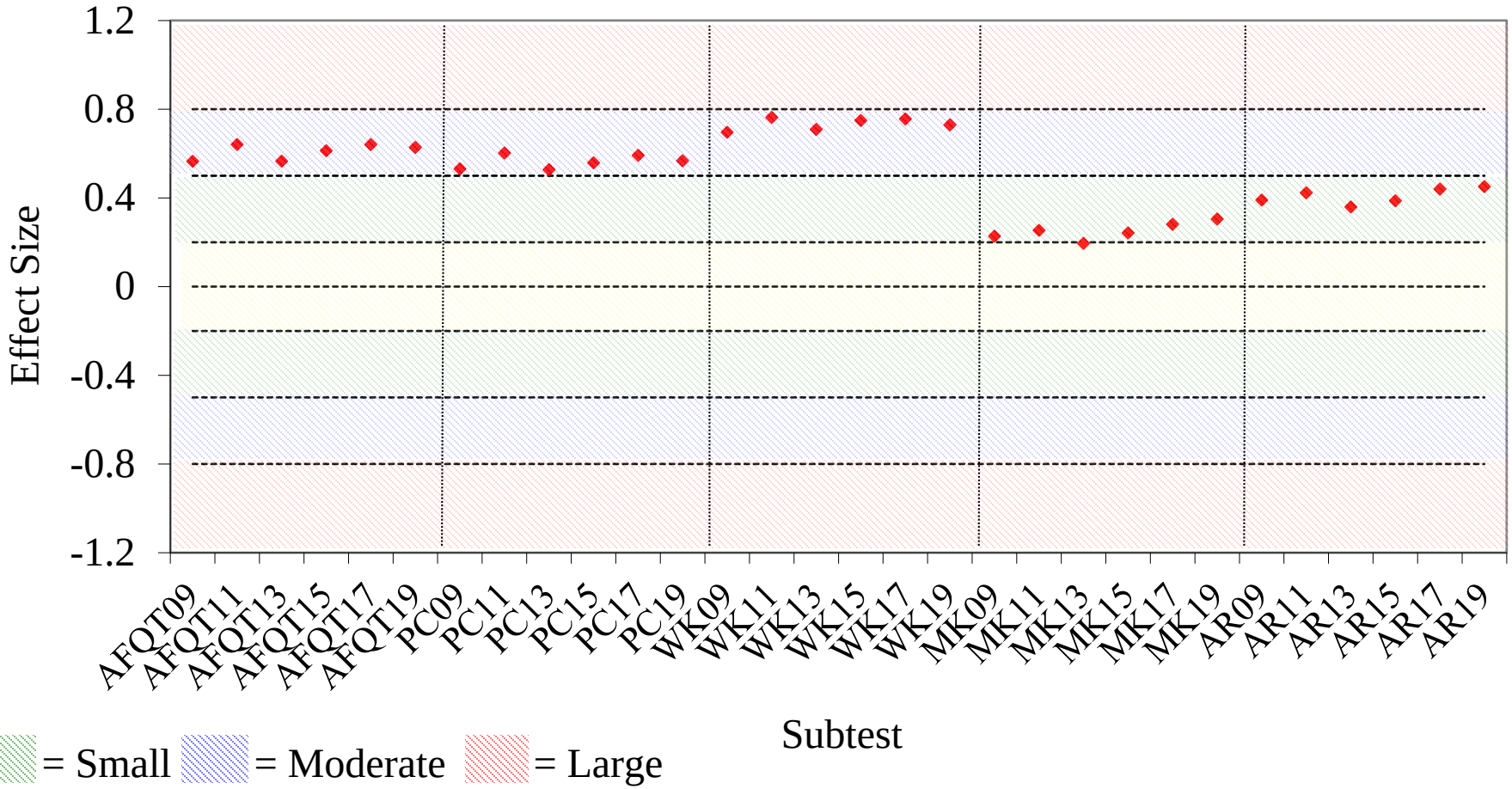
Comparison of Effect Sizes for FY09, FY11, FY13, FY15, FY17, FY19
Males Versus Females
Non-AFQT Tests



= Small = Moderate = Large

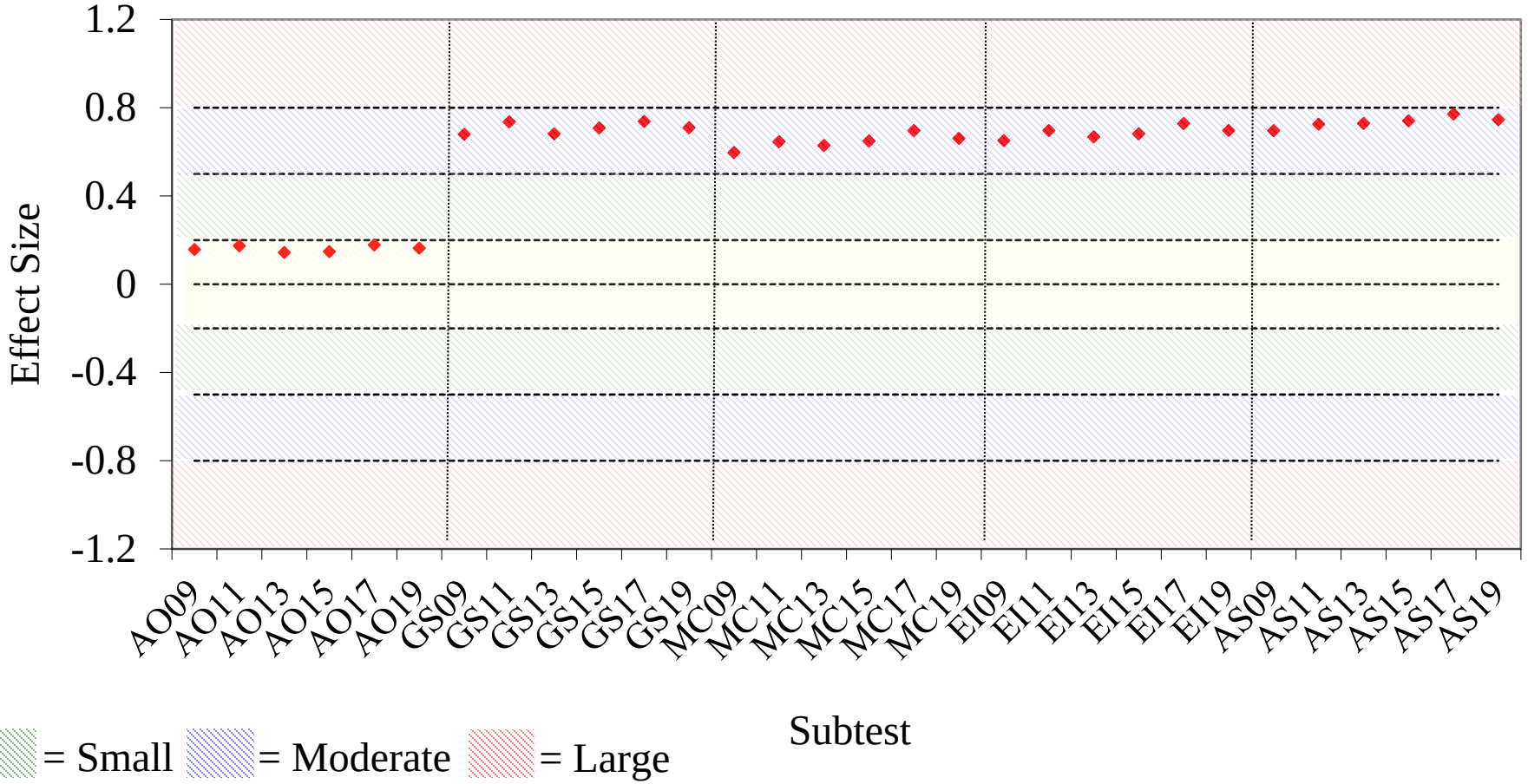
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes for FY09, FY11, FY13, FY15, FY17, FY19
Non-Hispanic Whites Versus Hispanic Whites
AFQT Tests/Scores



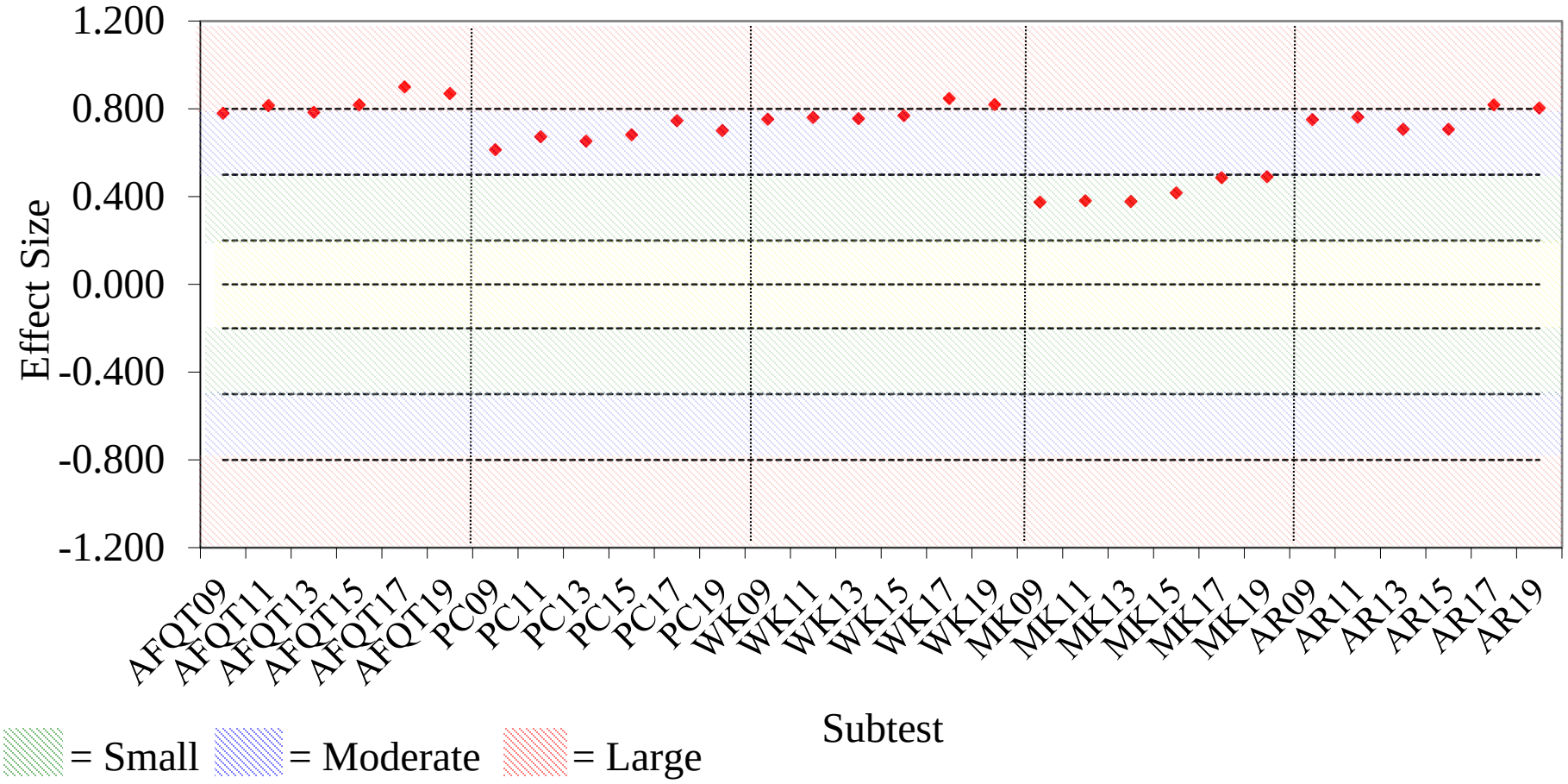
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes for FY09, FY11, FY13, FY15, FY17, FY19
Non-Hispanic Whites Versus Hispanic Whites
Non-AFQT Tests



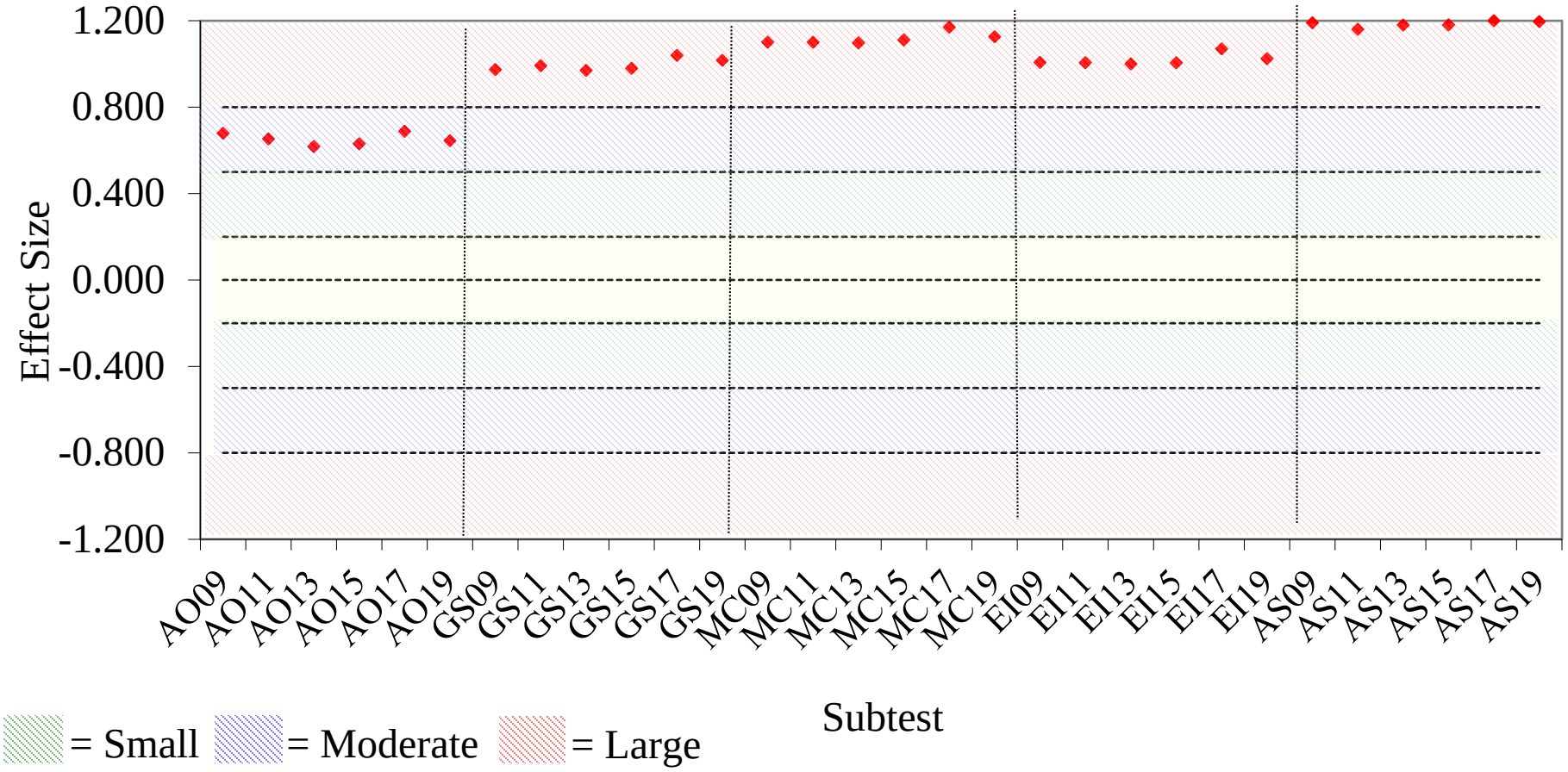
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes for FY09, FY11, FY13, FY15, FY17, FY19
Non-Hispanic Whites Versus Non-Hispanic Blacks
AFQT Tests/Scores



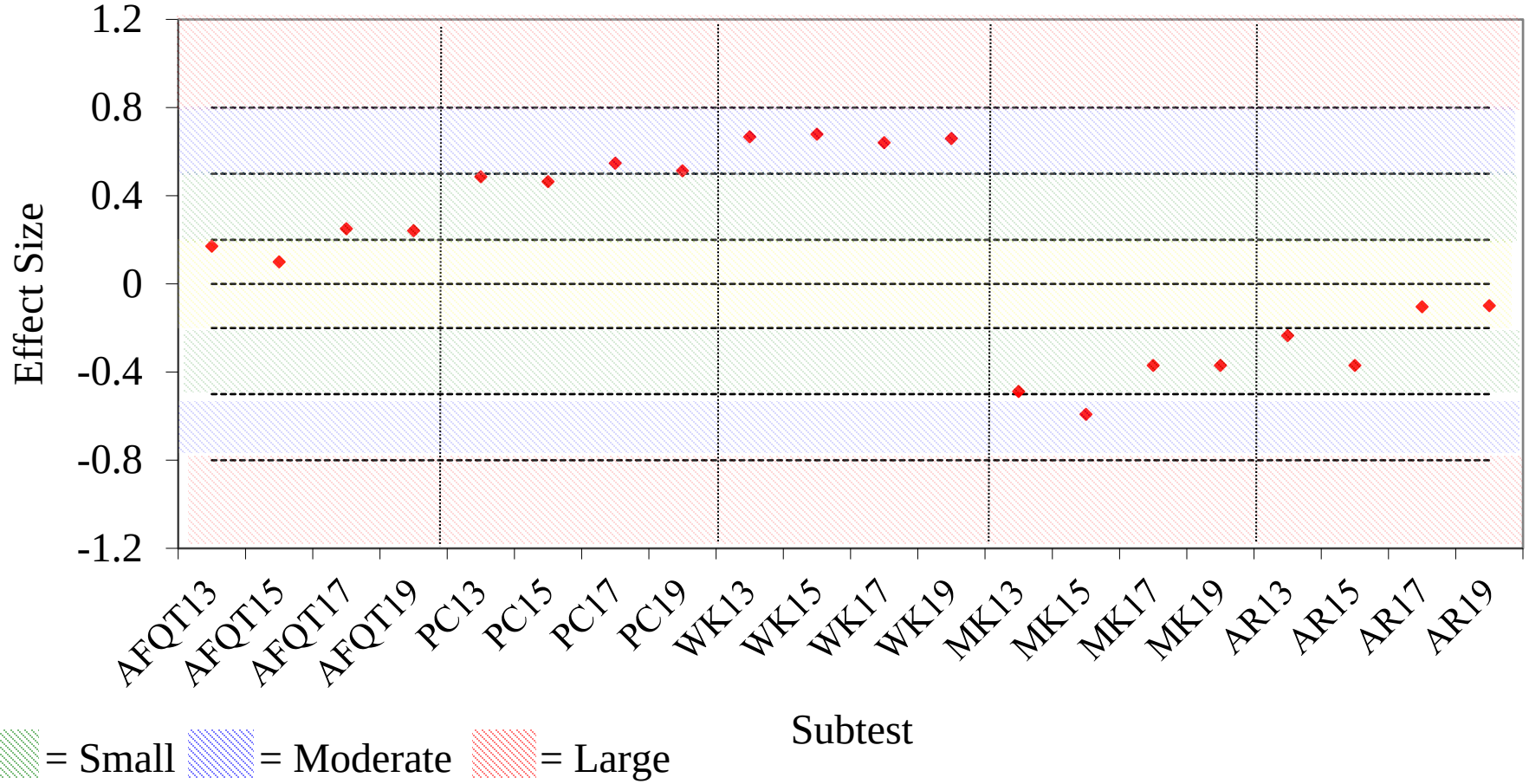
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes for FY09, FY11, FY13, FY15, FY17, FY19
Non-Hispanic Whites Versus Non-Hispanic Blacks
Non-AFQT Tests



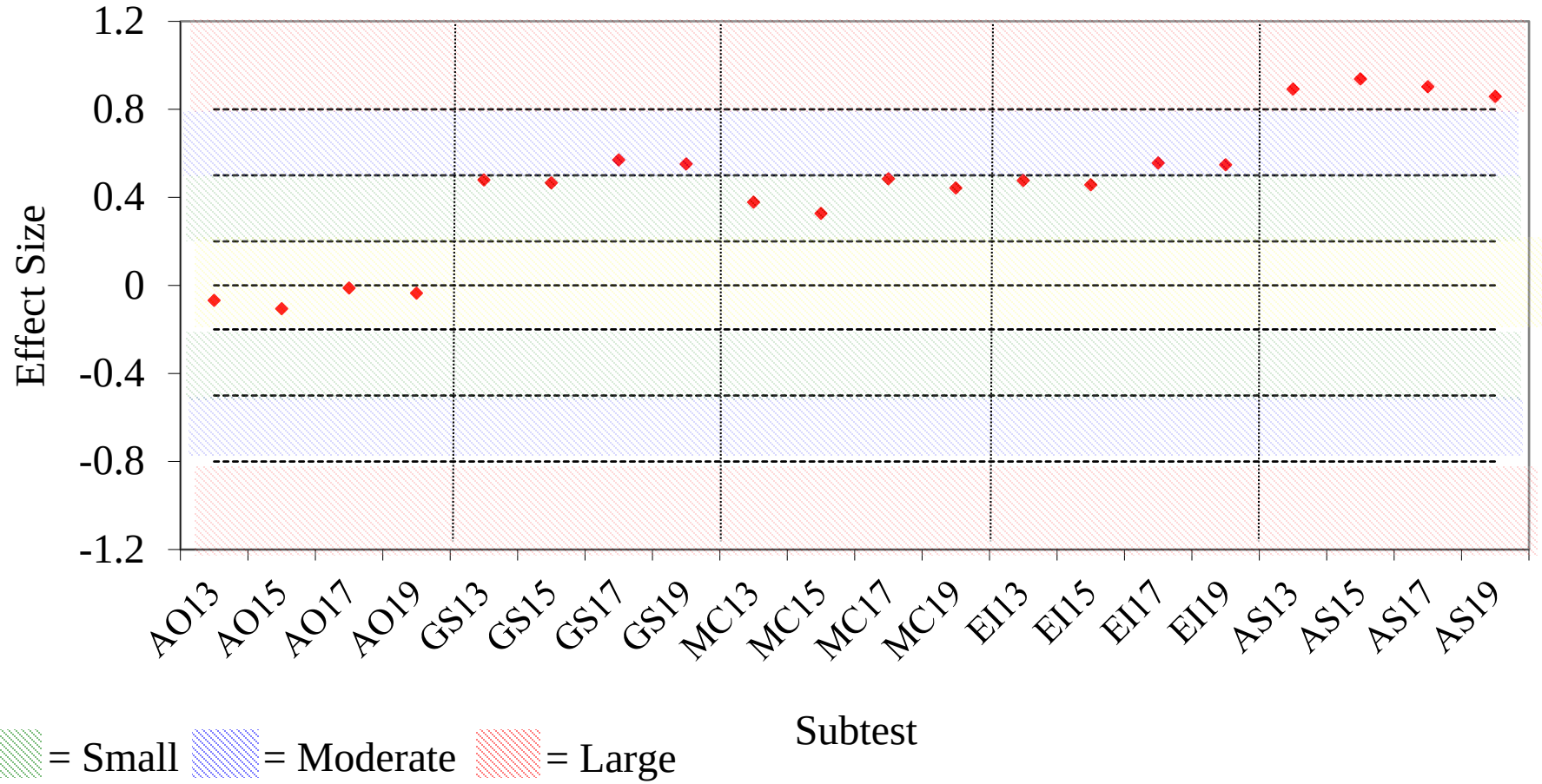
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes for FY13, FY15, FY17, FY19
Non-Hispanic Whites Versus Non-Hispanic Asians
AFQT Tests/Scores



OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes for FY13, FY15, FY17, FY19
Non-Hispanic Whites Versus Non-Hispanic Asians
Non-AFQT Tests

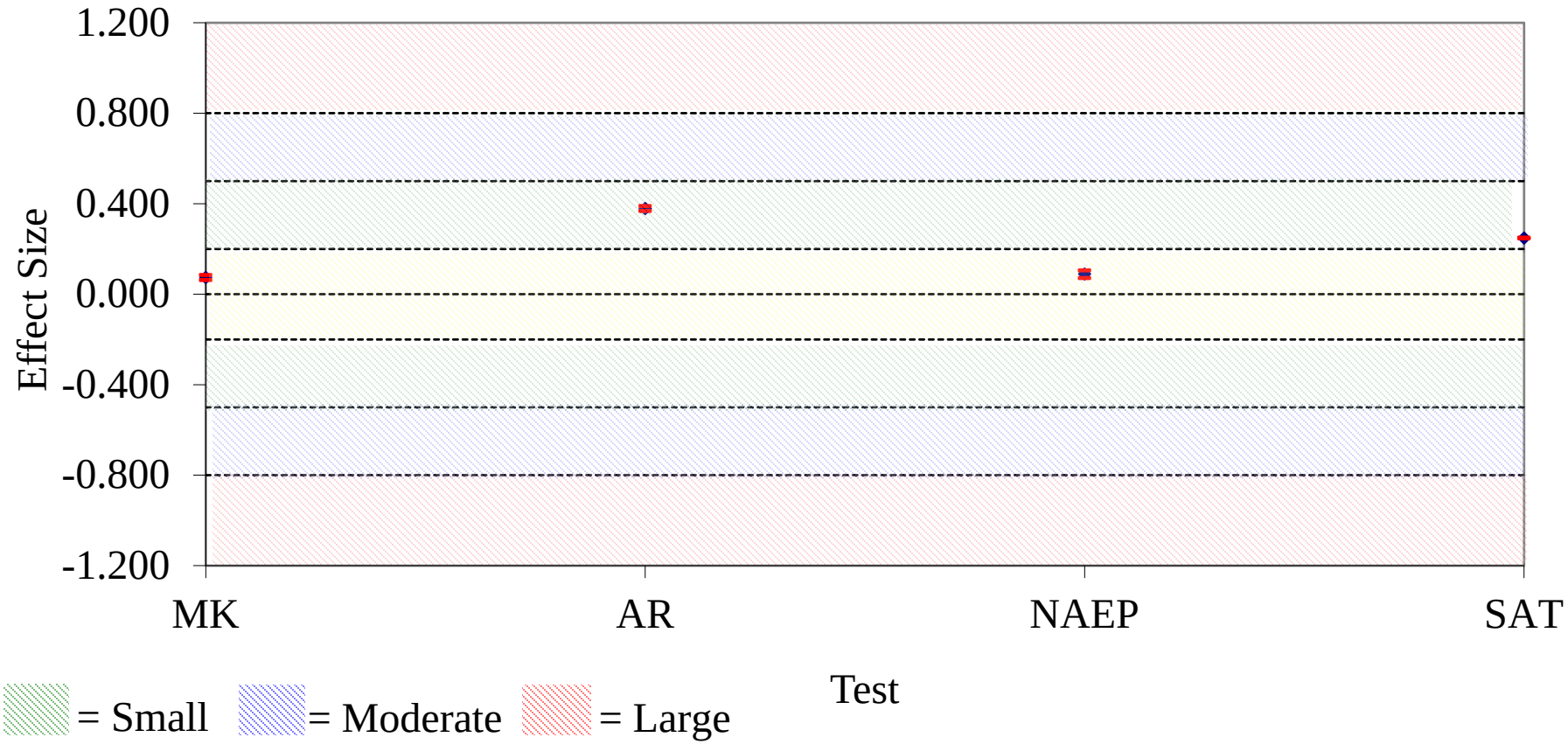


OPA
OFFICE OF PEOPLE ANALYTICS

WHAT DOES IT MEAN?

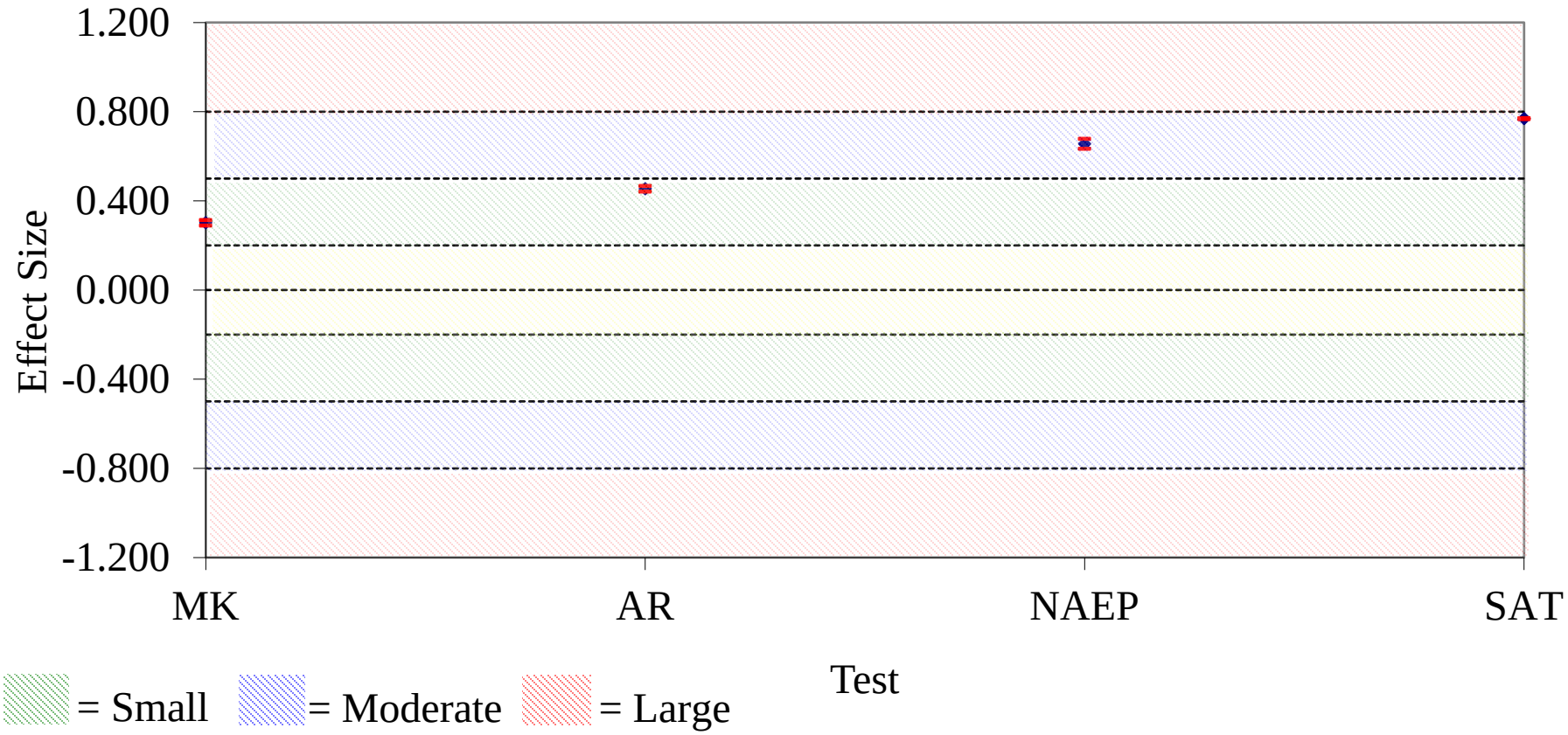
- The magnitude of impact on the ASVAB has remained fairly constant across fiscal years, but still varies in size from negligible to large across tests and groups.
- A comparison of impact across different testing programs gives some indication of whether the observed FY2019 magnitudes are reasonable.
- 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 – 2015 Grade 12 (Reading, Math, and Science)

Comparison of Effect Sizes Across Testing Programs
Content Area = Math
Males Versus Females



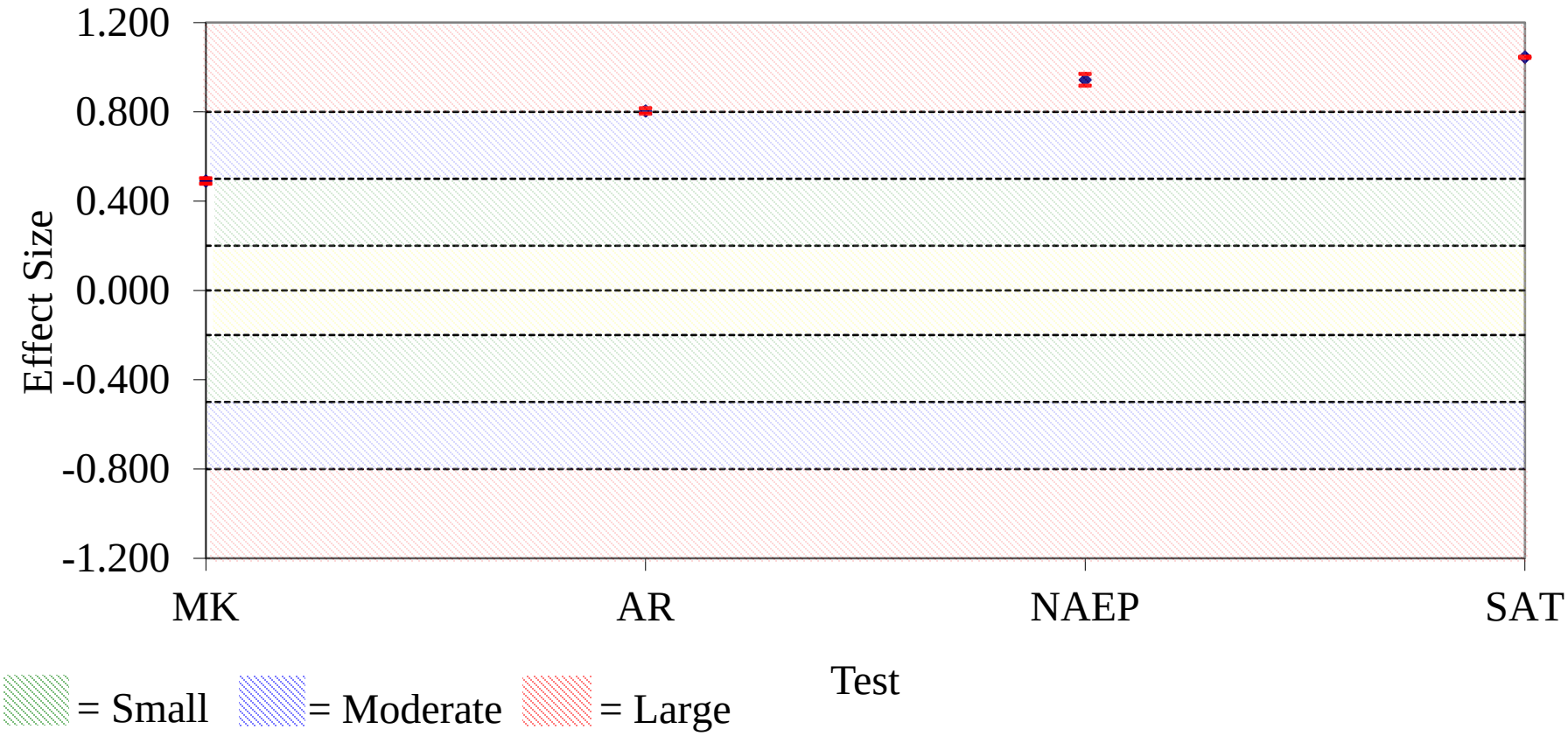
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Math
Non-Hispanic Whites Versus Hispanics



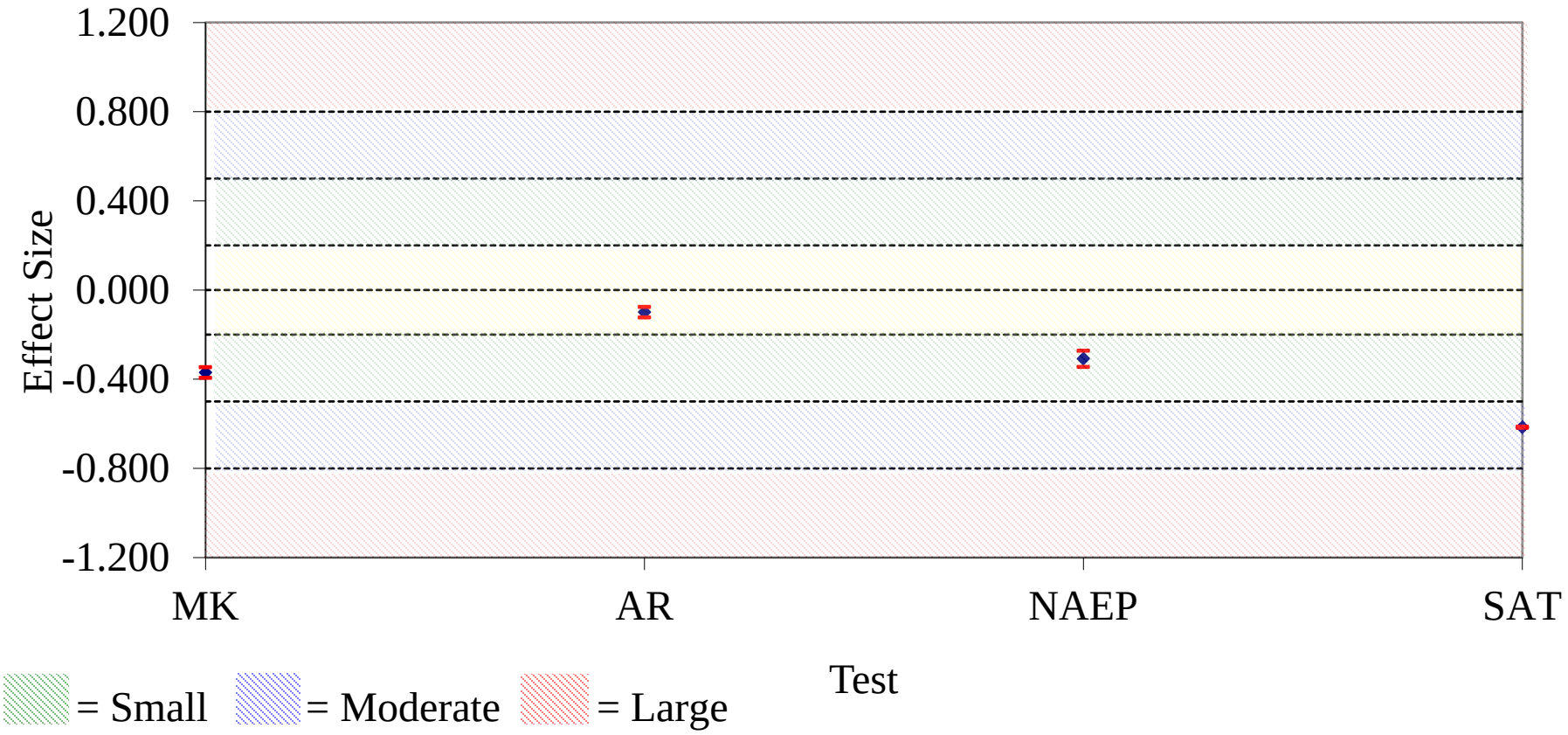
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Math
Non-Hispanic Whites Versus Non-Hispanic Blacks



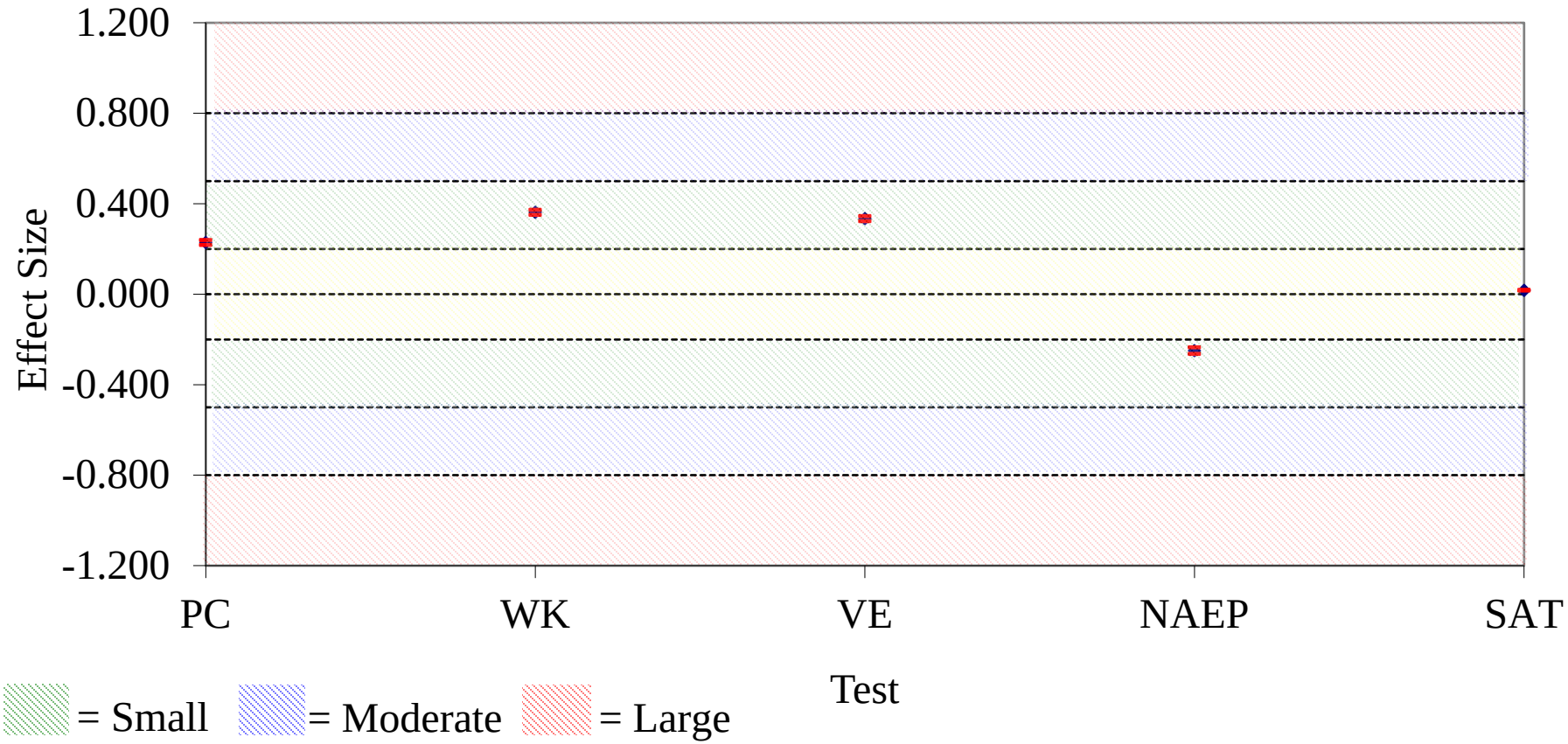
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Math
Non-Hispanic Whites Versus Asians



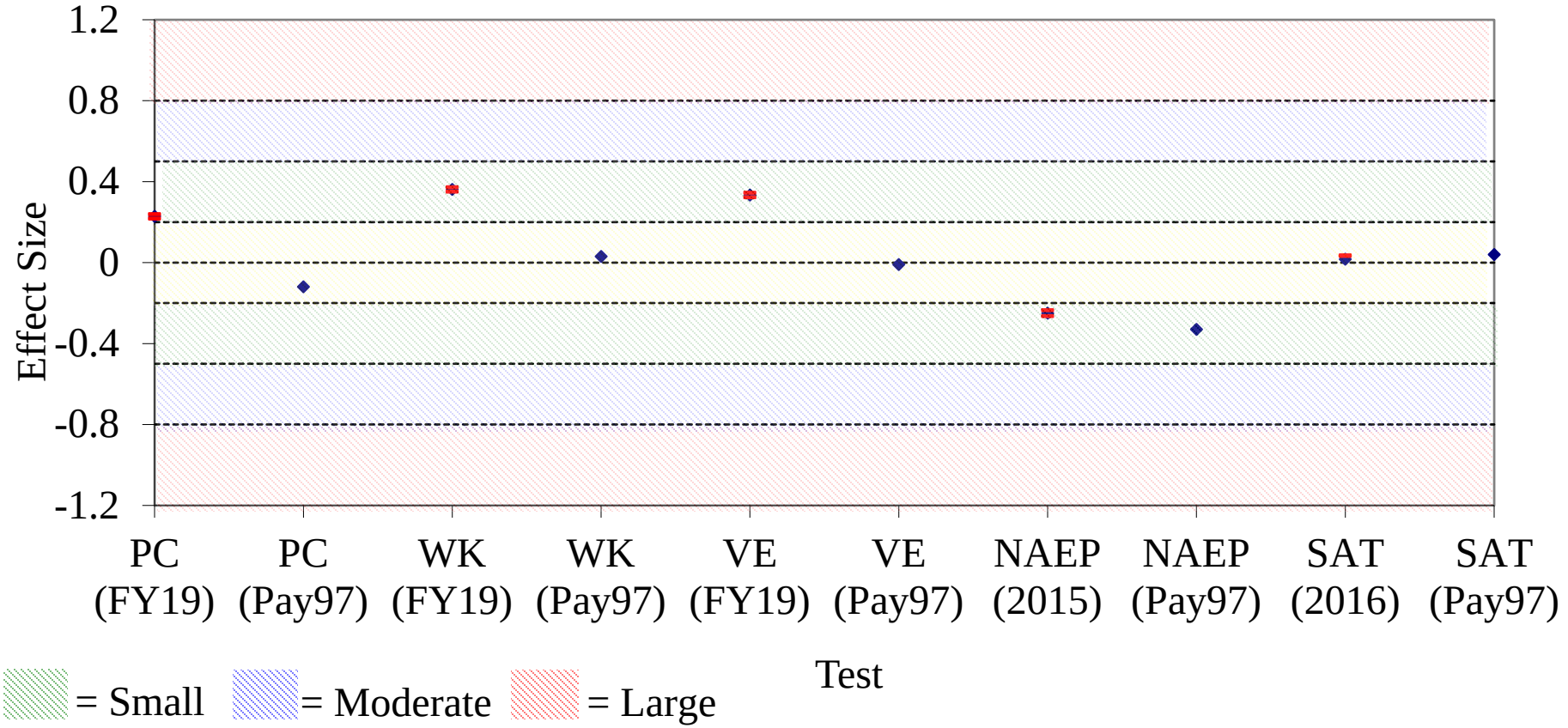
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Reading/Verbal
Males Versus Females



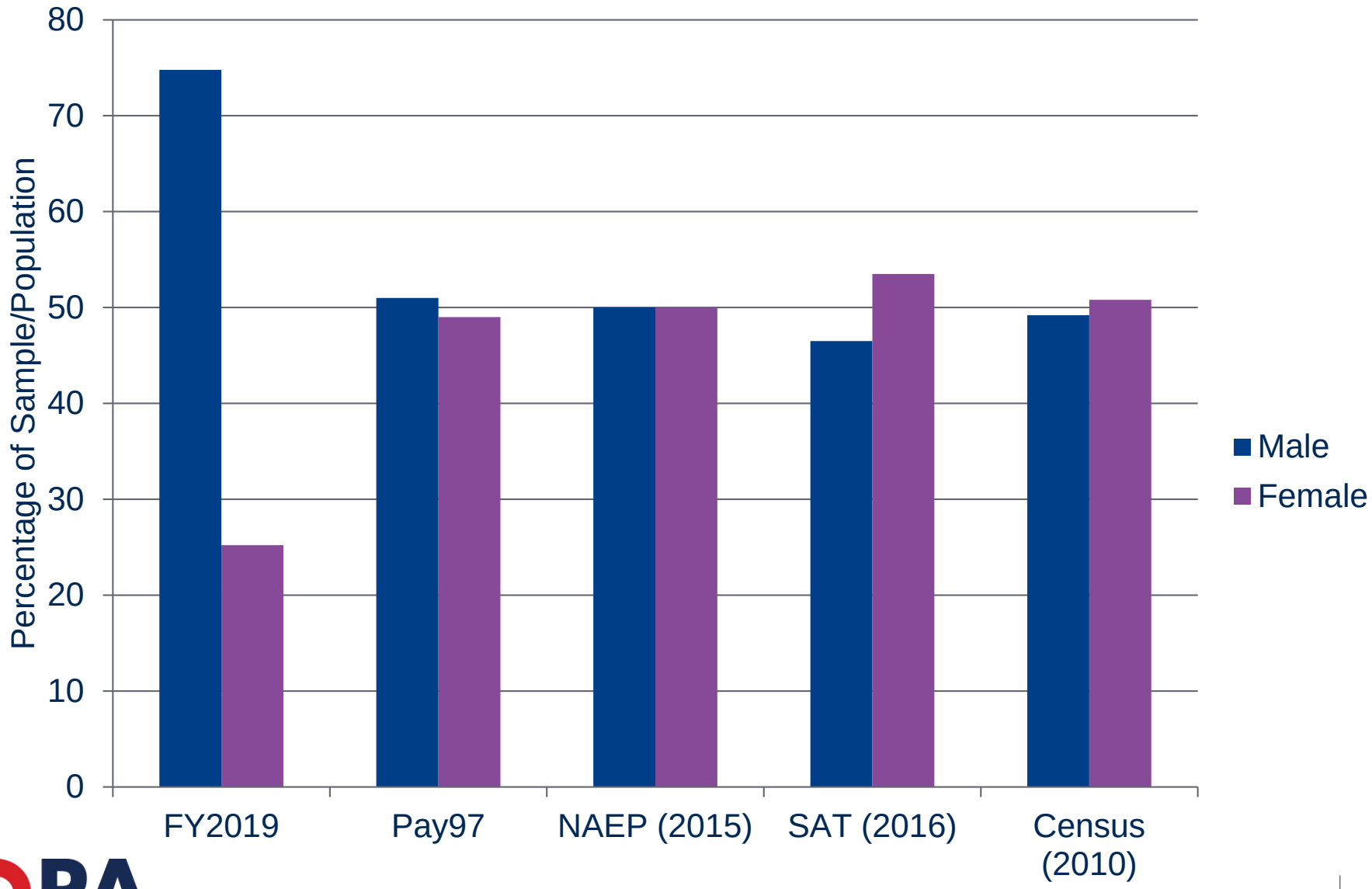
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Reading/Verbal
Males Versus Females



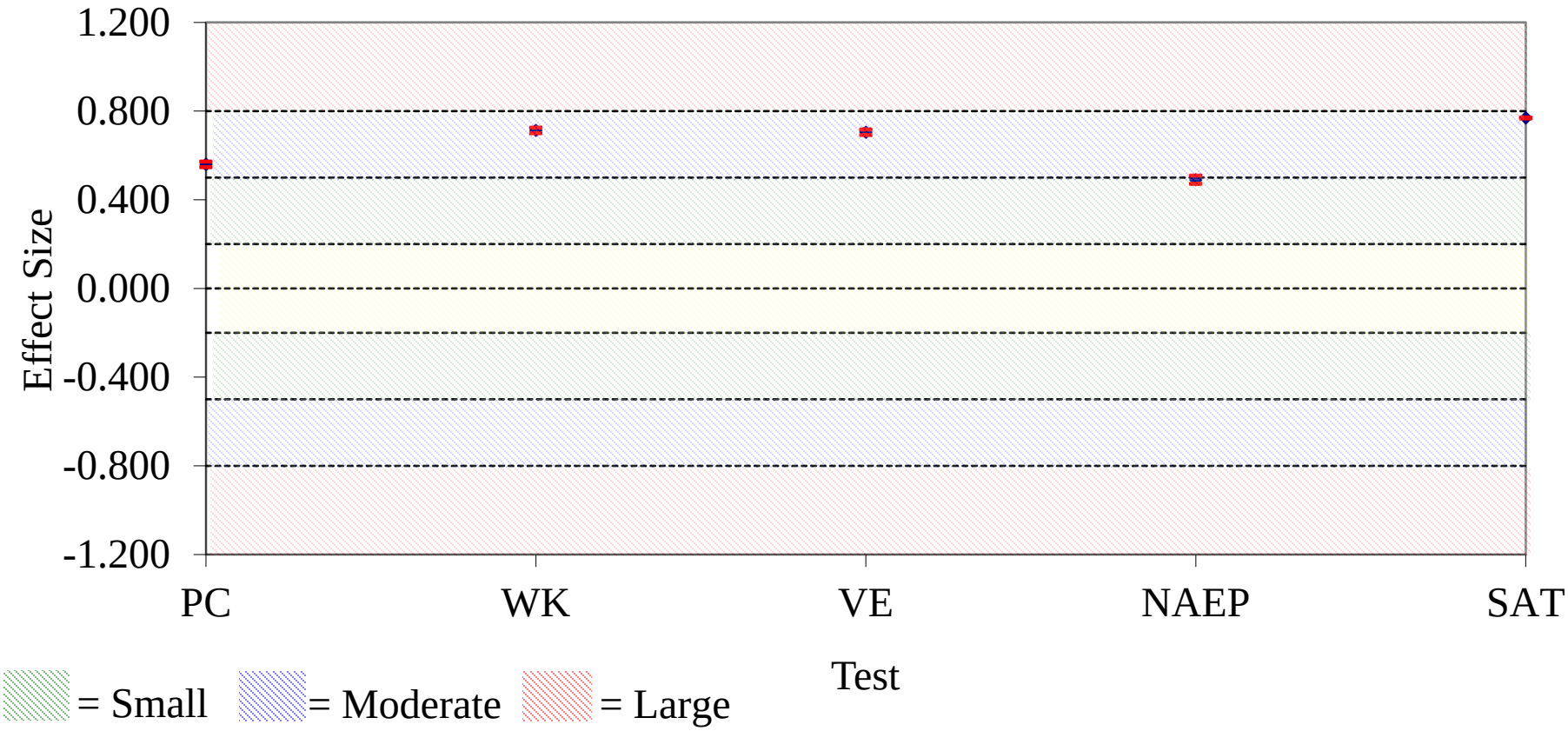
OPA
OFFICE OF PEOPLE ANALYTICS

Gender Representation Across Samples/Populations



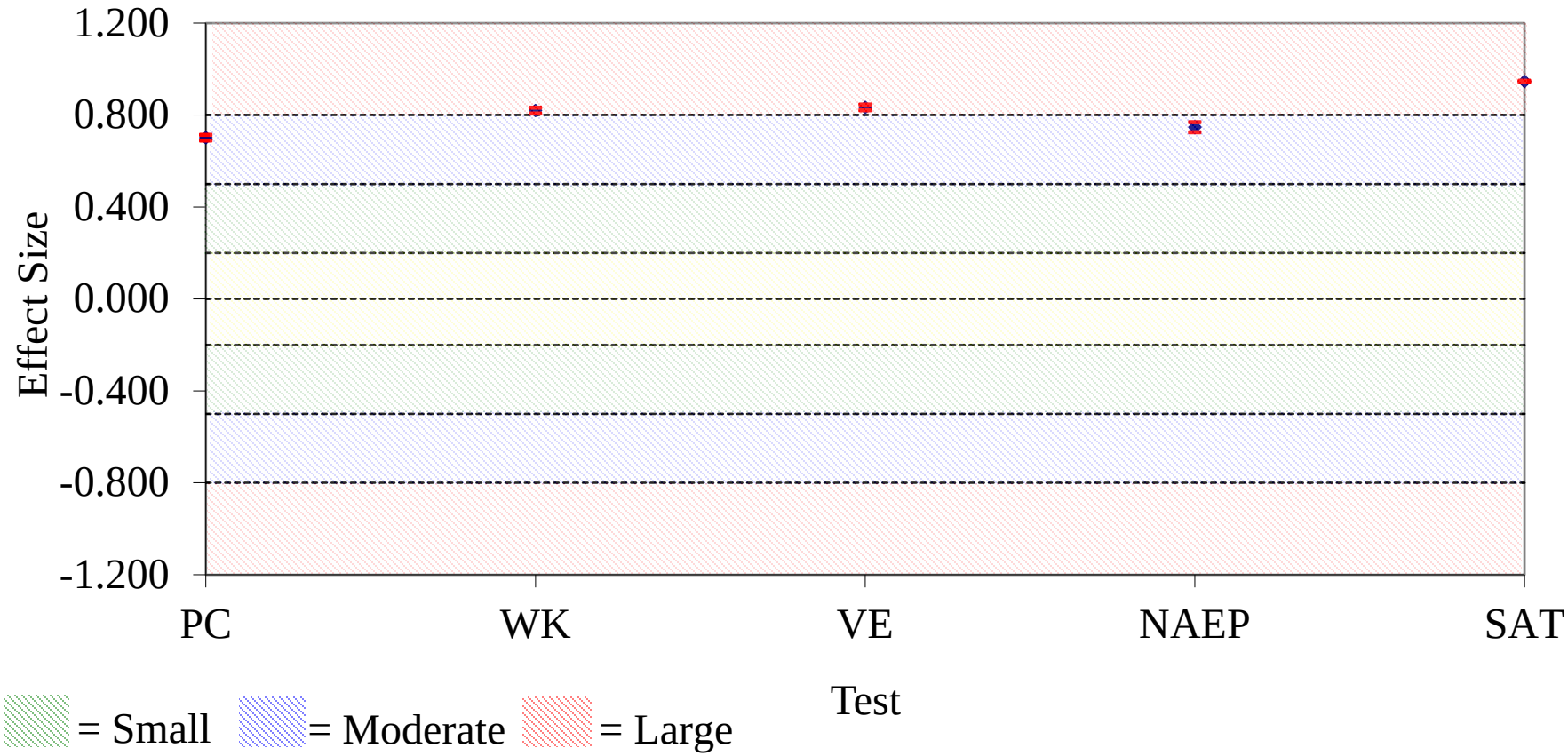
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Reading/Verbal
Non-Hispanic Whites Versus Hispanics



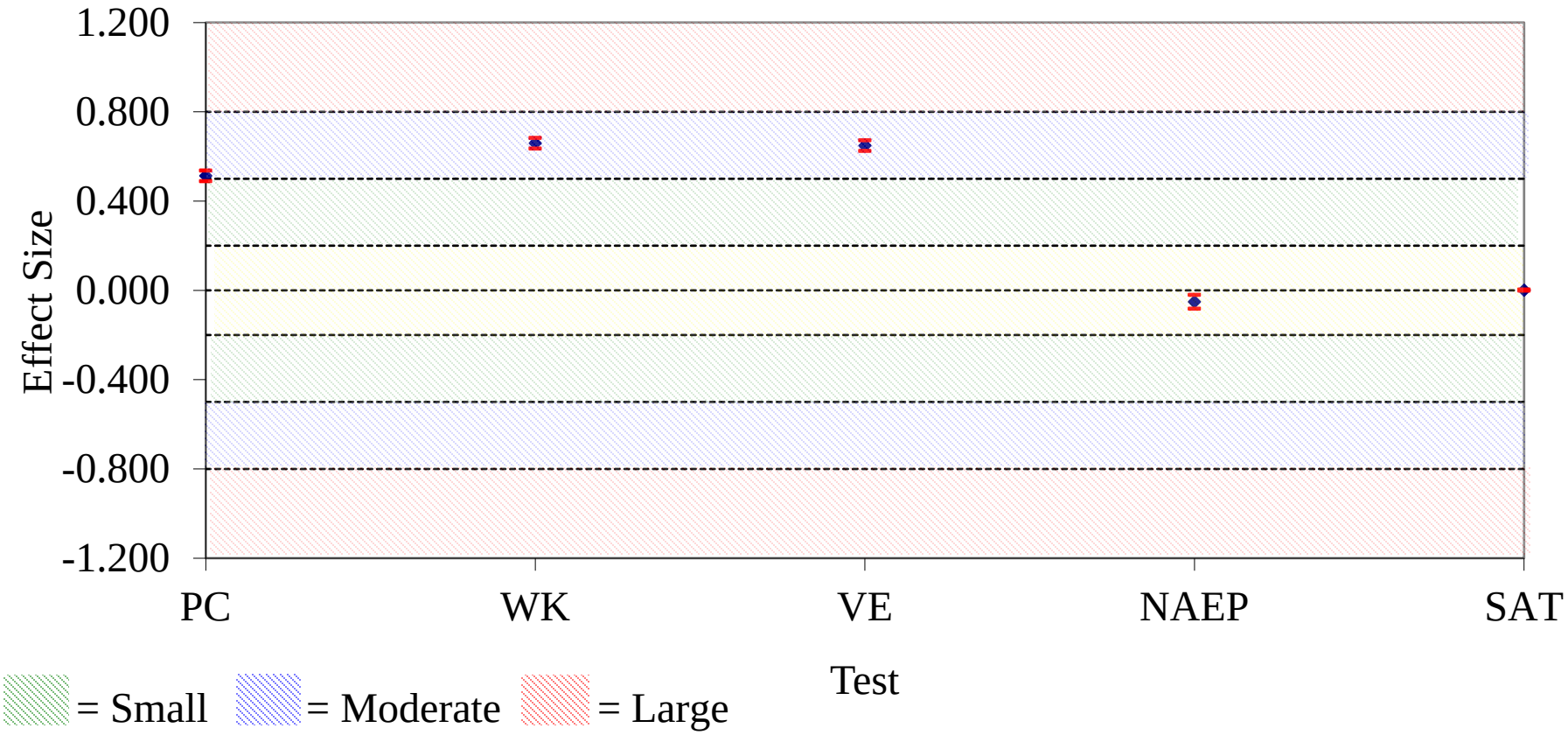
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Reading/Verbal
Non-Hispanic Whites Versus Non-Hispanic Blacks



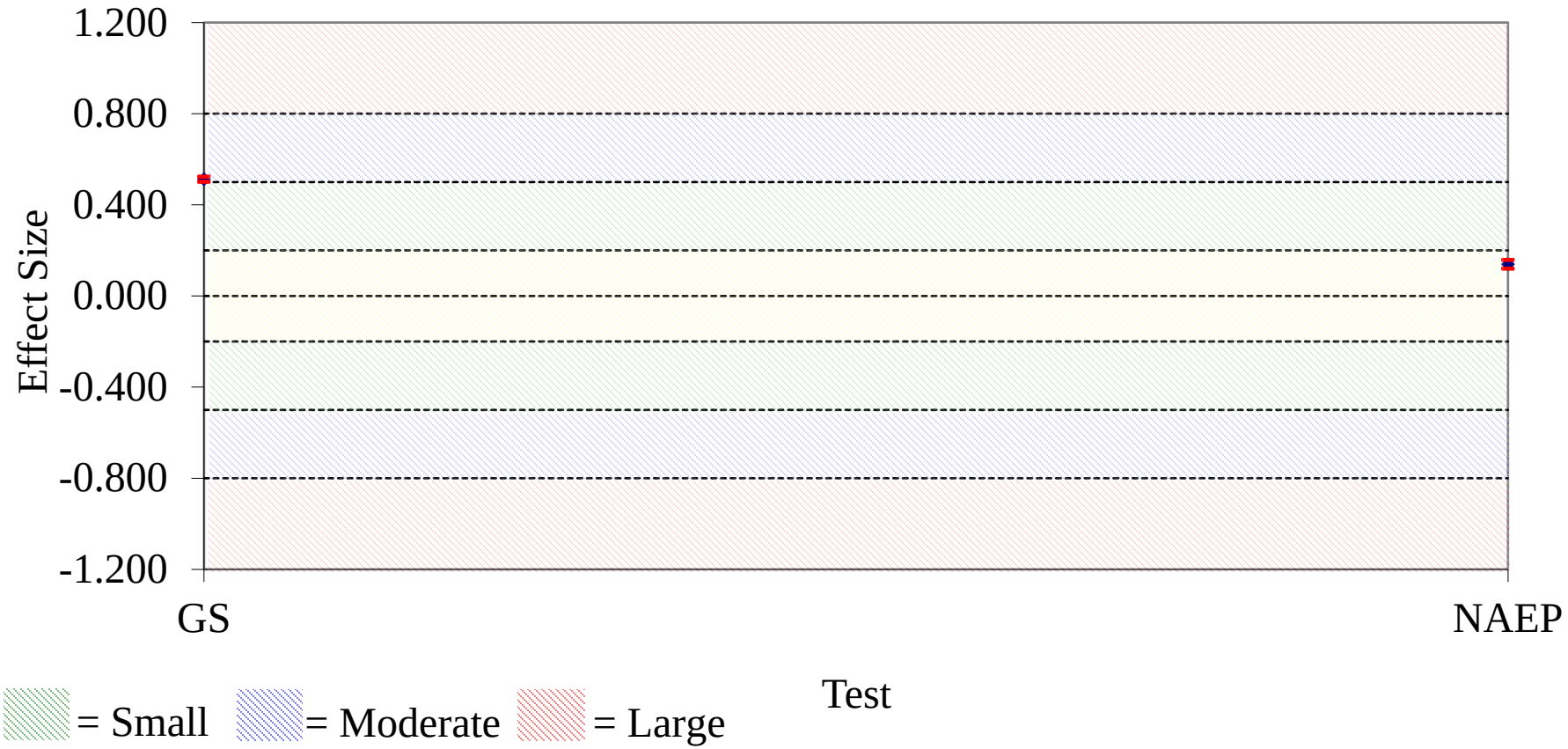
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Reading/Verbal
Non-Hispanic Whites Versus Asians



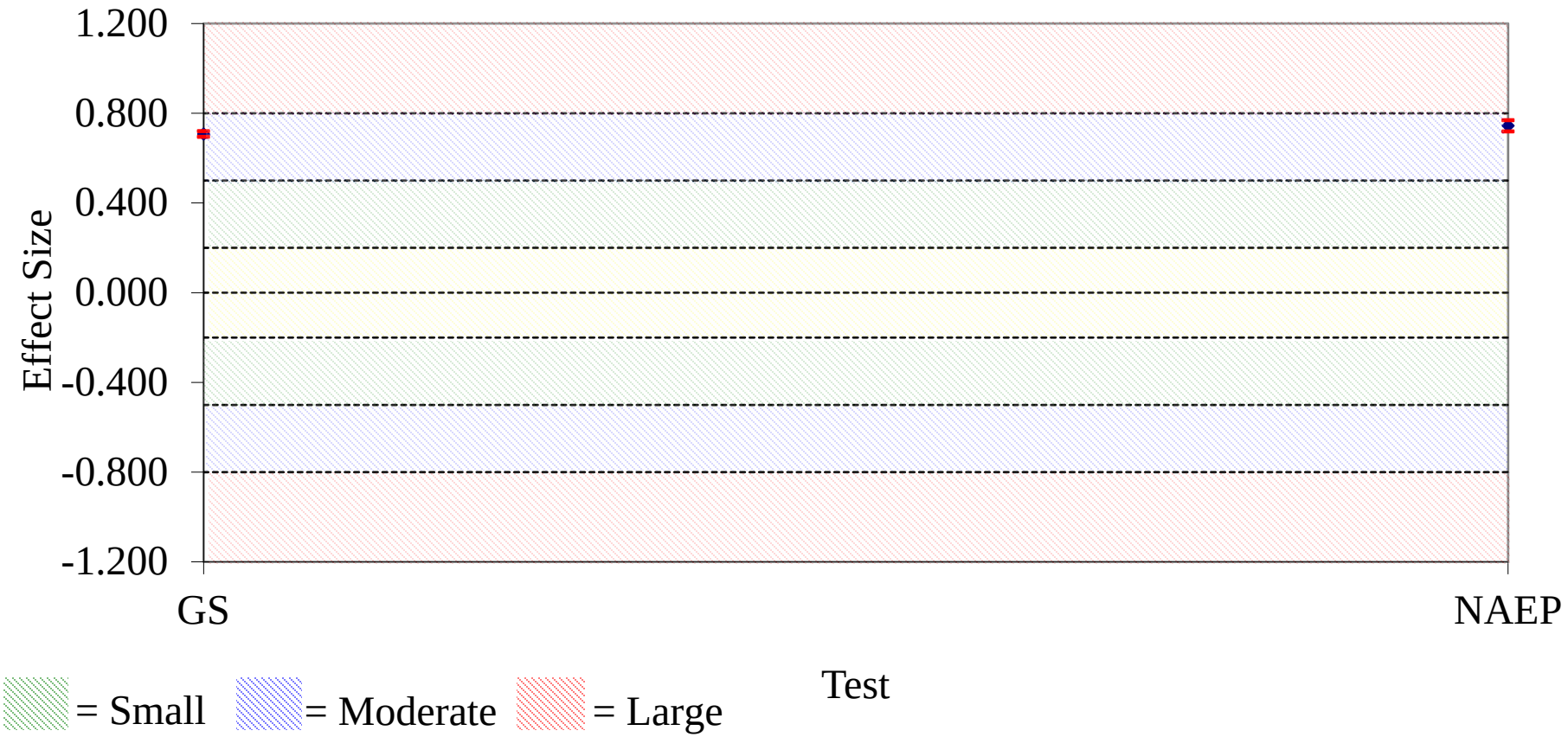
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Science
Males Versus Females



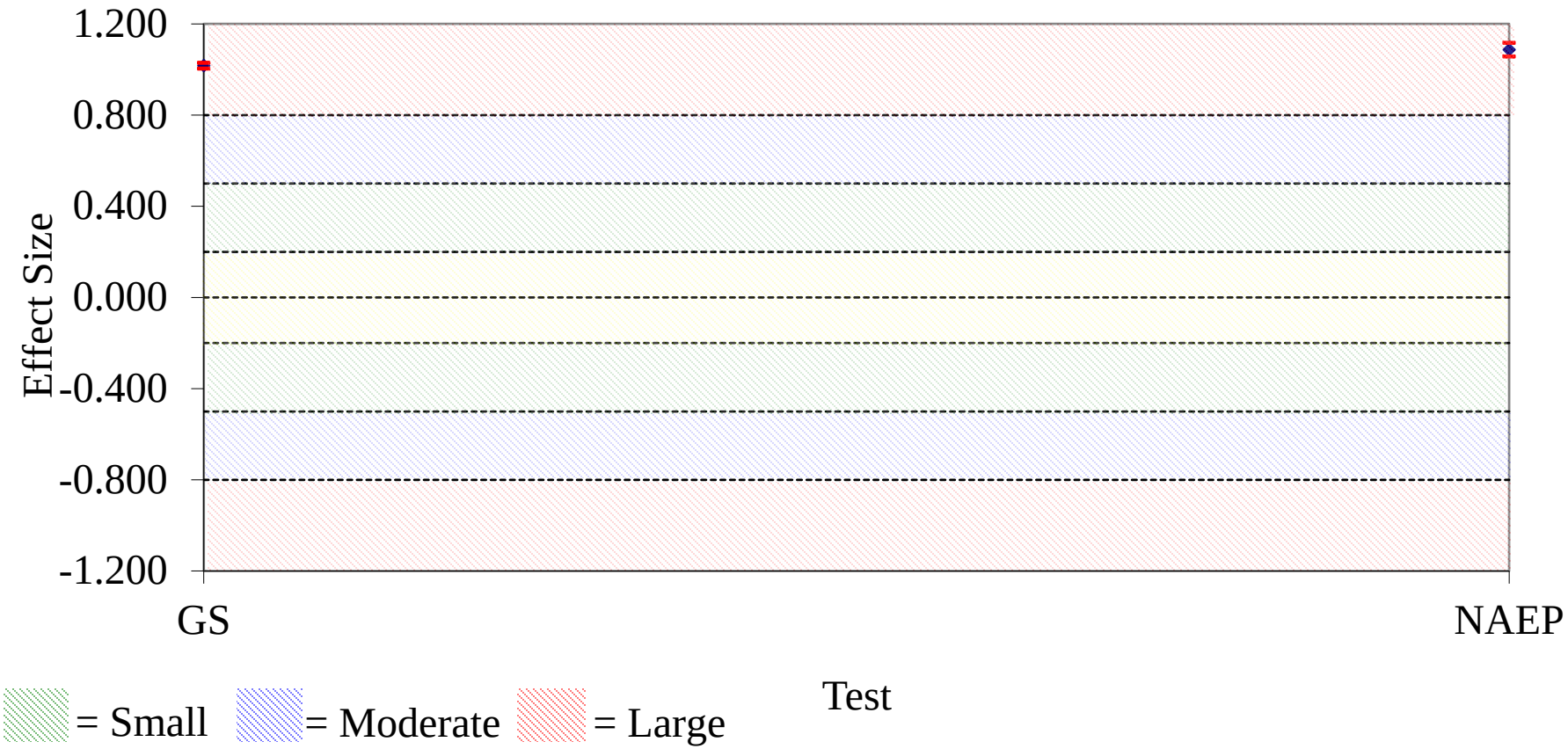
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Science
Non-Hispanic Whites Versus Hispanics



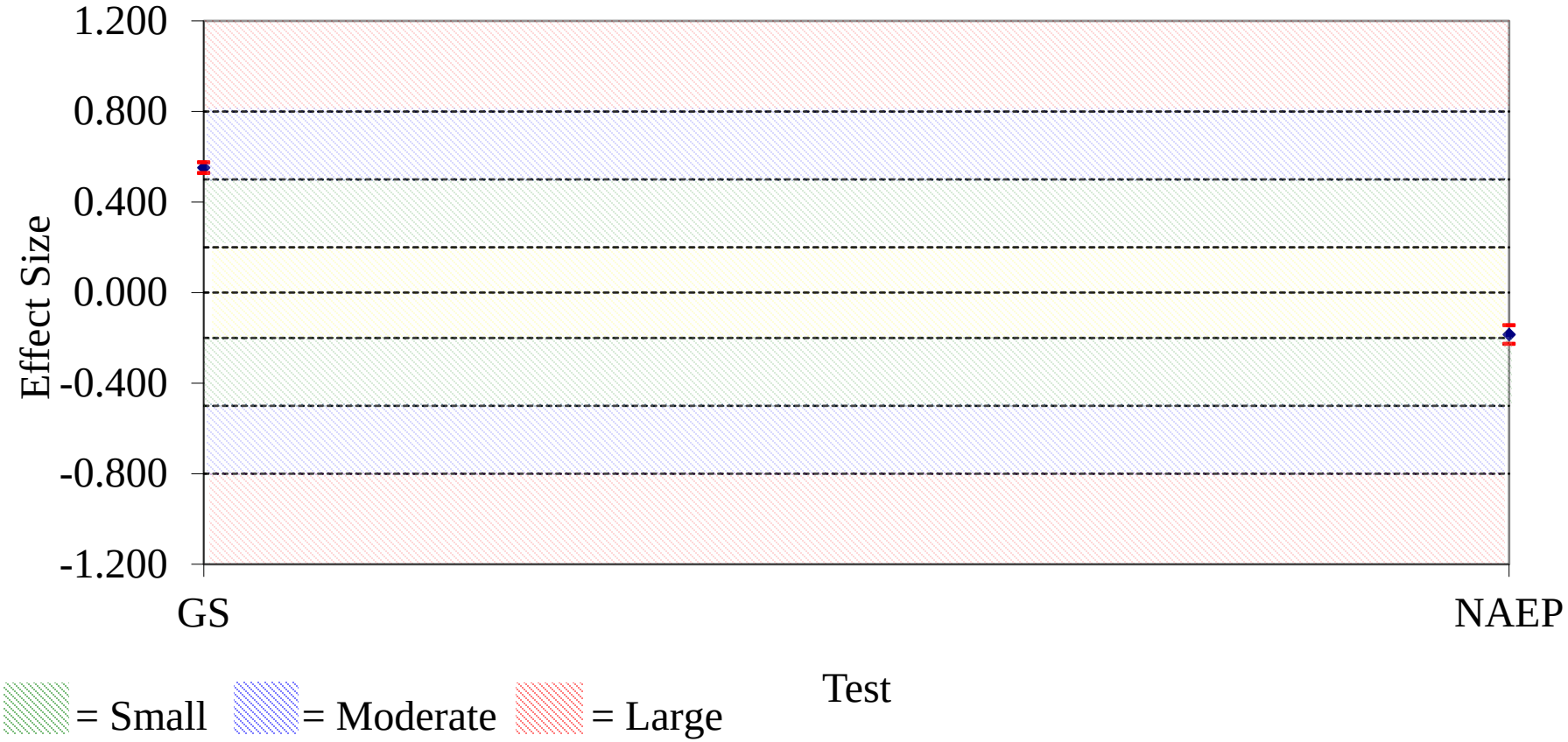
OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Science
Non-Hispanic Whites Versus Non-Hispanic Blacks



OPA
OFFICE OF PEOPLE ANALYTICS

Comparison of Effect Sizes Across Testing Programs
Content Area = Science
Non-Hispanic Whites Versus Asians



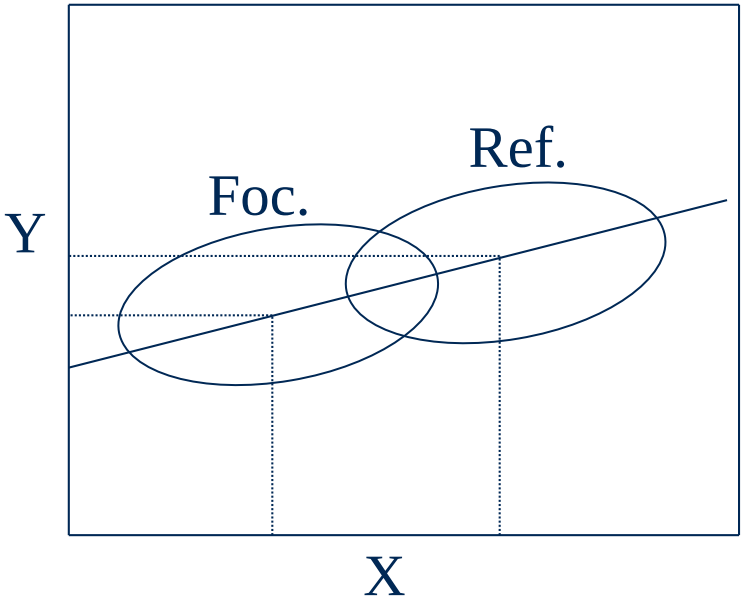
OPA
OFFICE OF PEOPLE ANALYTICS

CONCLUSIONS AND CAVEATS

- For the AFQT tests (and GS), the direction and magnitude of overall impact is generally consistent with that observed on comparable SAT and NAEP tests, which suggests that impact on ASVAB tests may reflect legitimate differences in the studied groups.
  - Comparisons across programs may be somewhat restricted due to differences in group definitions, testing populations, test content, etc.
- "To the extent that members of one group do more poorly on a subtest of items that are a *legitimate part of the content domain*, we would be reluctant to call the discrepancy evidence of *bias*" (Shepard, 1987).

CONCLUSIONS AND CAVEATS

- Adverse impact does *not* reflect 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 bias is indicated when the regression line relating the test score [X] and a criterion [Y] is the same for each group.



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

CONCLUSIONS AND CAVEATS

- Previous research on the ASVAB technical tests showed similar prediction lines across (1) males and females and (2) blacks and whites (Wise et al., 1992), suggesting no bias for the tests and groups studied.
  - DMDC recommended in 2010 that an updated validity study be conducted for relevant tests and groups.
  - Lack of access to criterion data across Services (except Air Force) presents an impediment to updating the study.
  - More recent thinking in the realm of bias detection is that regression-based approaches may not accurately reflect bias.
- Recent acquisition of training outcome data from the Services may make it possible to examine AFQT for test bias.

CONCLUSIONS AND CAVEATS

- Moderated Multiple Regression (for FSG) and Logistic Regression (for P/F) may be used to evaluate intercept and slop bias for AFQT scores and group membership when prediction training outcomes.
  - Lack of variance on the P/F criterion and relatively rare outcome presents a challenge (mostly pass)
- Better to look to the future? Reducing potential impact will be a high priority when considering revisions to the ASVAB and AFQT contents.

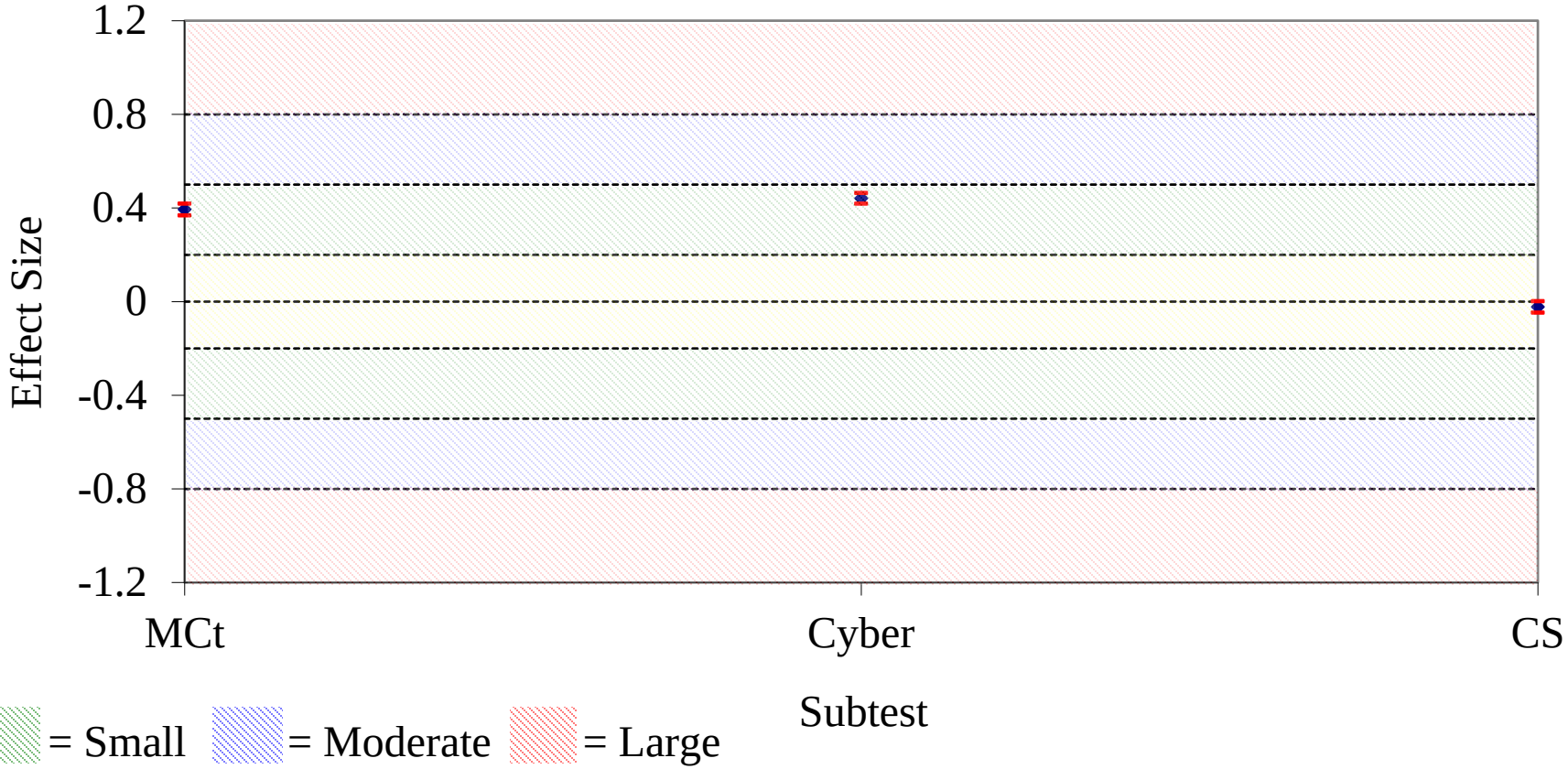
SPECIAL TESTS ON THE ASVAB PLATFORM

- **Mental Counters (MCt)**: A counting test of working memory (Navy only)

- **Cyber Test (Cyber):** Test of basic computer and information systems knowledge (All Services)

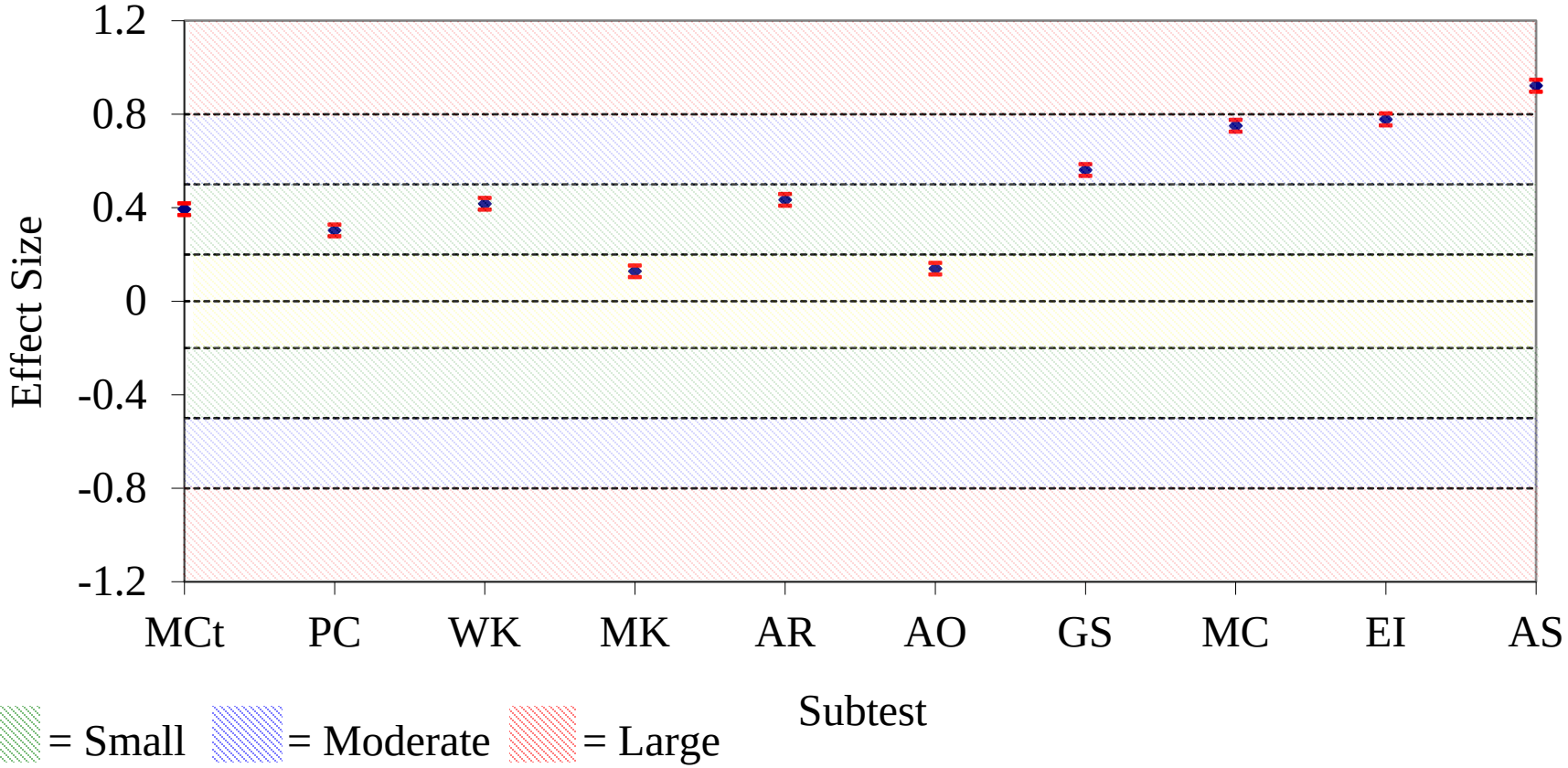
- **Coding Speed (CS):** A speeded test of assigning code numbers to words (Navy only)

Effect Sizes (and 95% Confidence Interval) for Special Tests Scores
Males v. Females
FY2019



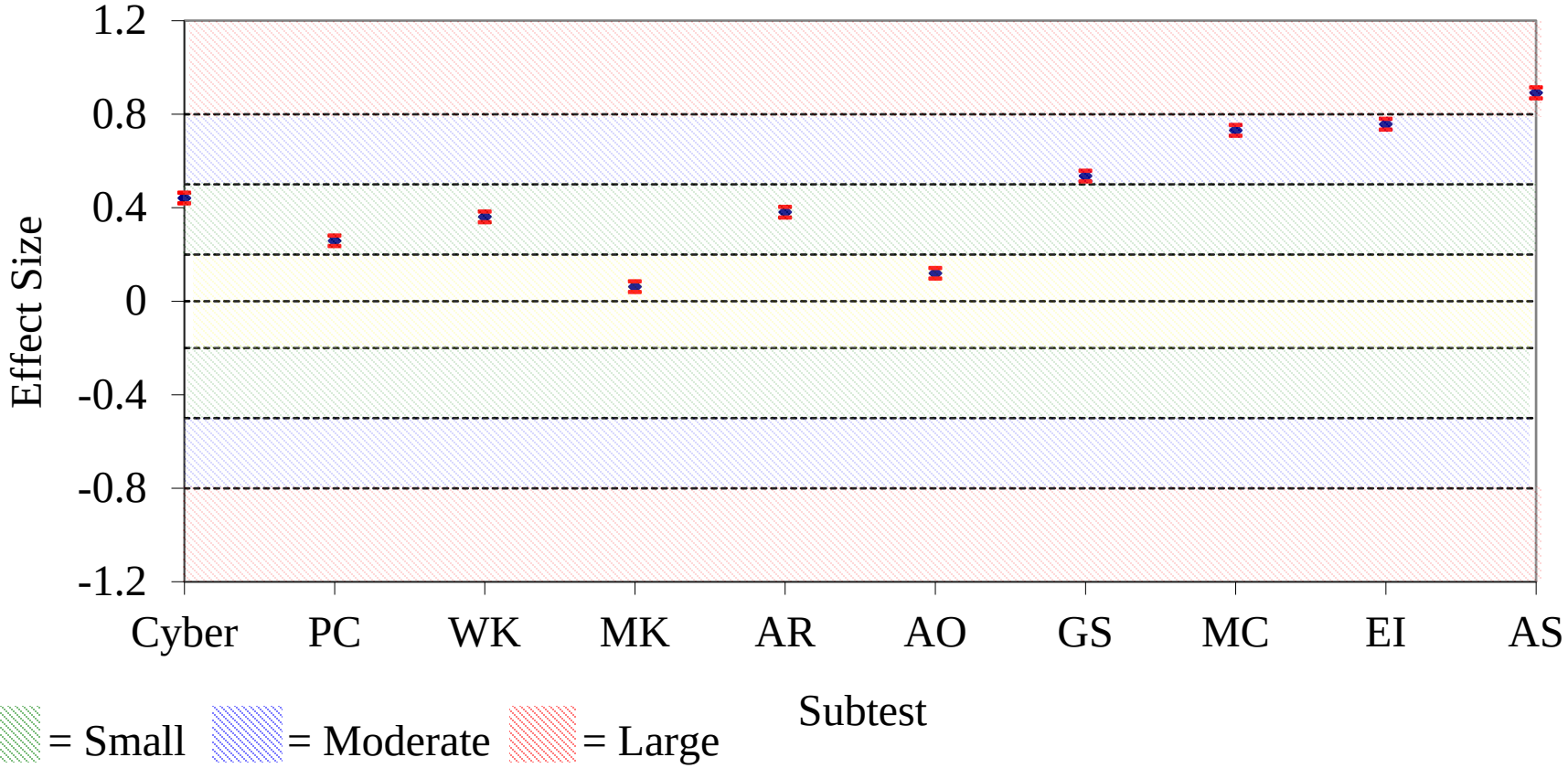
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Males Versus Females, MCt Sample
FY2019



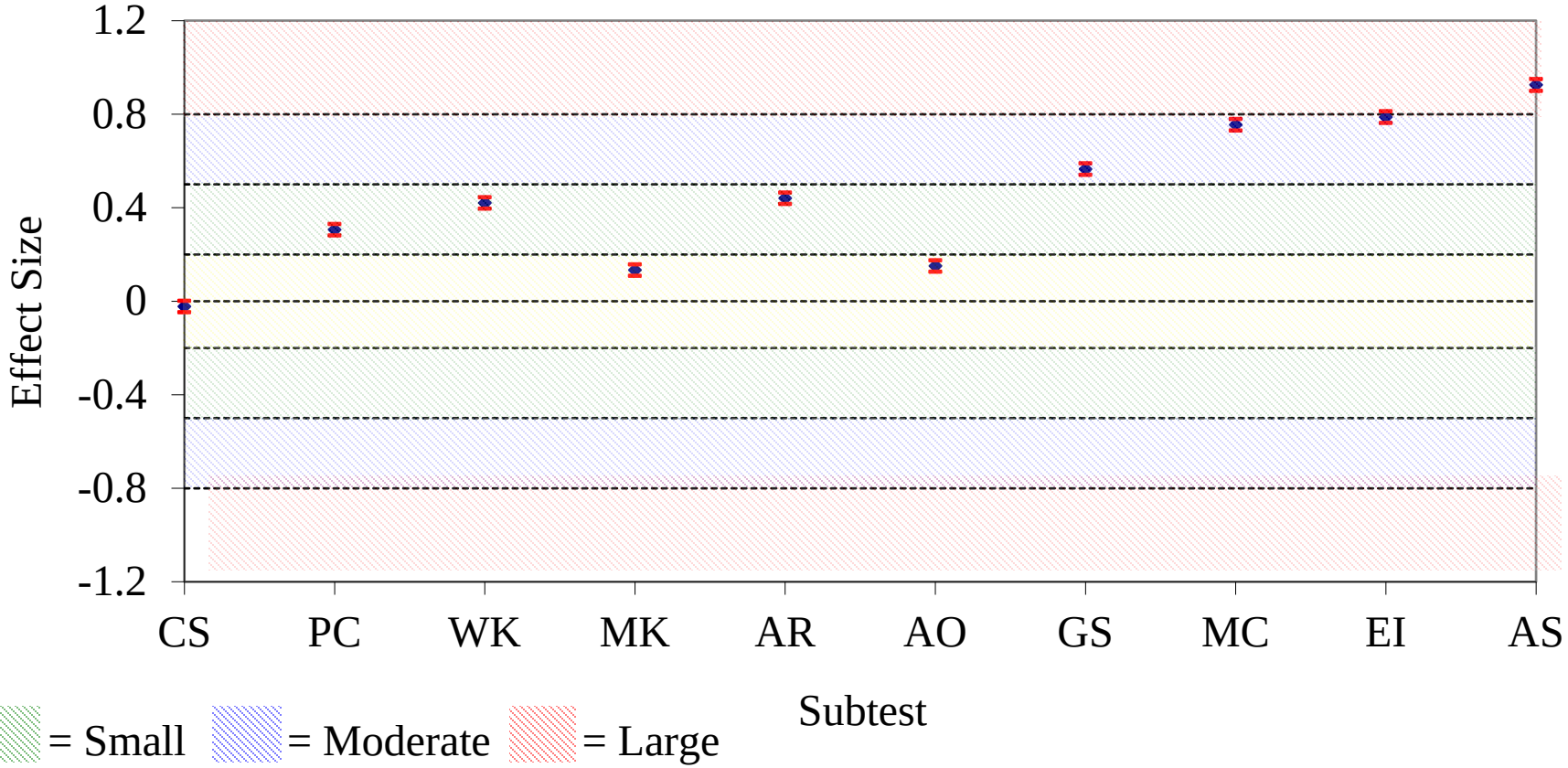
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Males Versus Females, Cyber Sample
FY2019



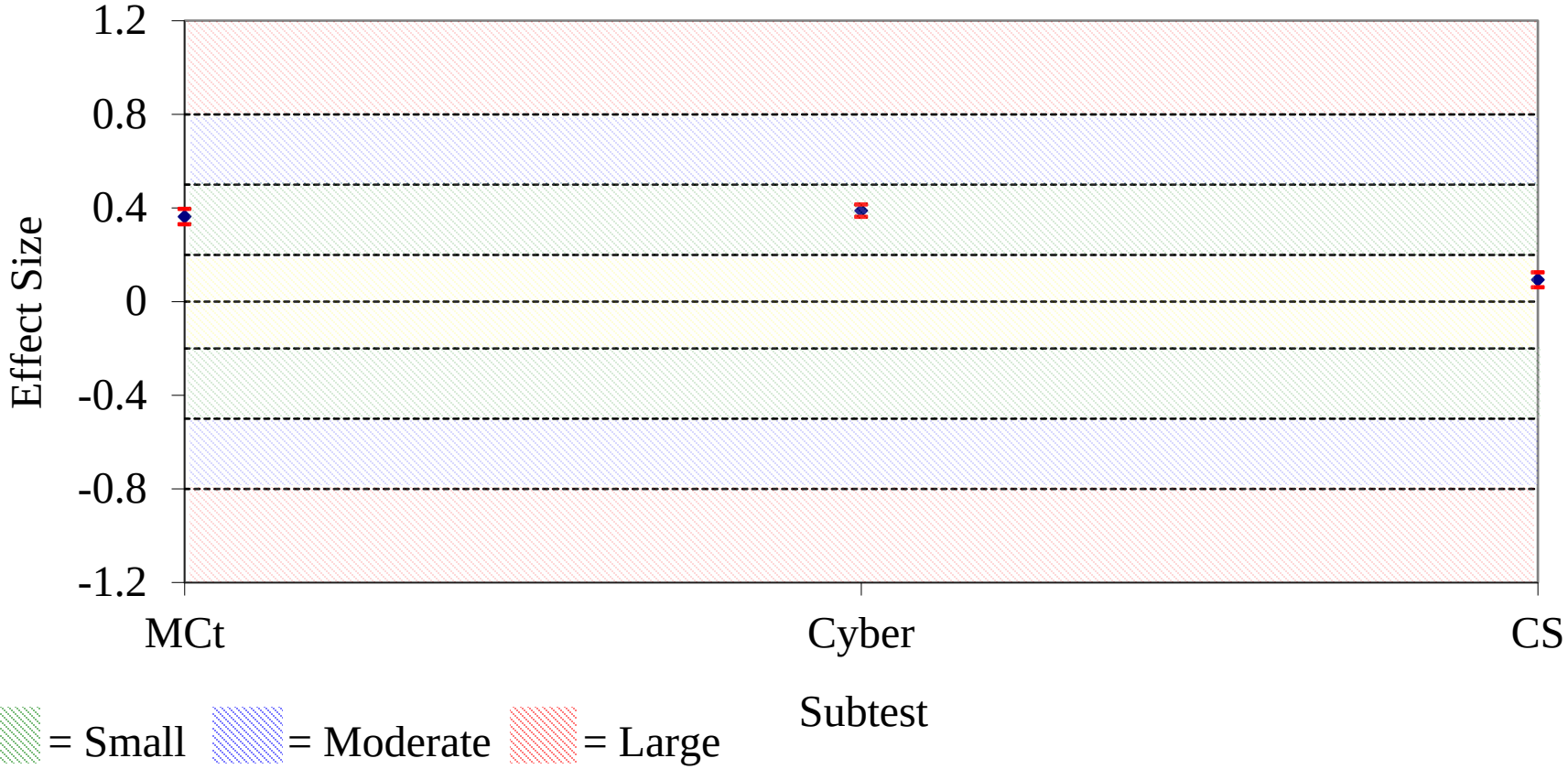
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Males Versus Females, CS Sample
FY2019



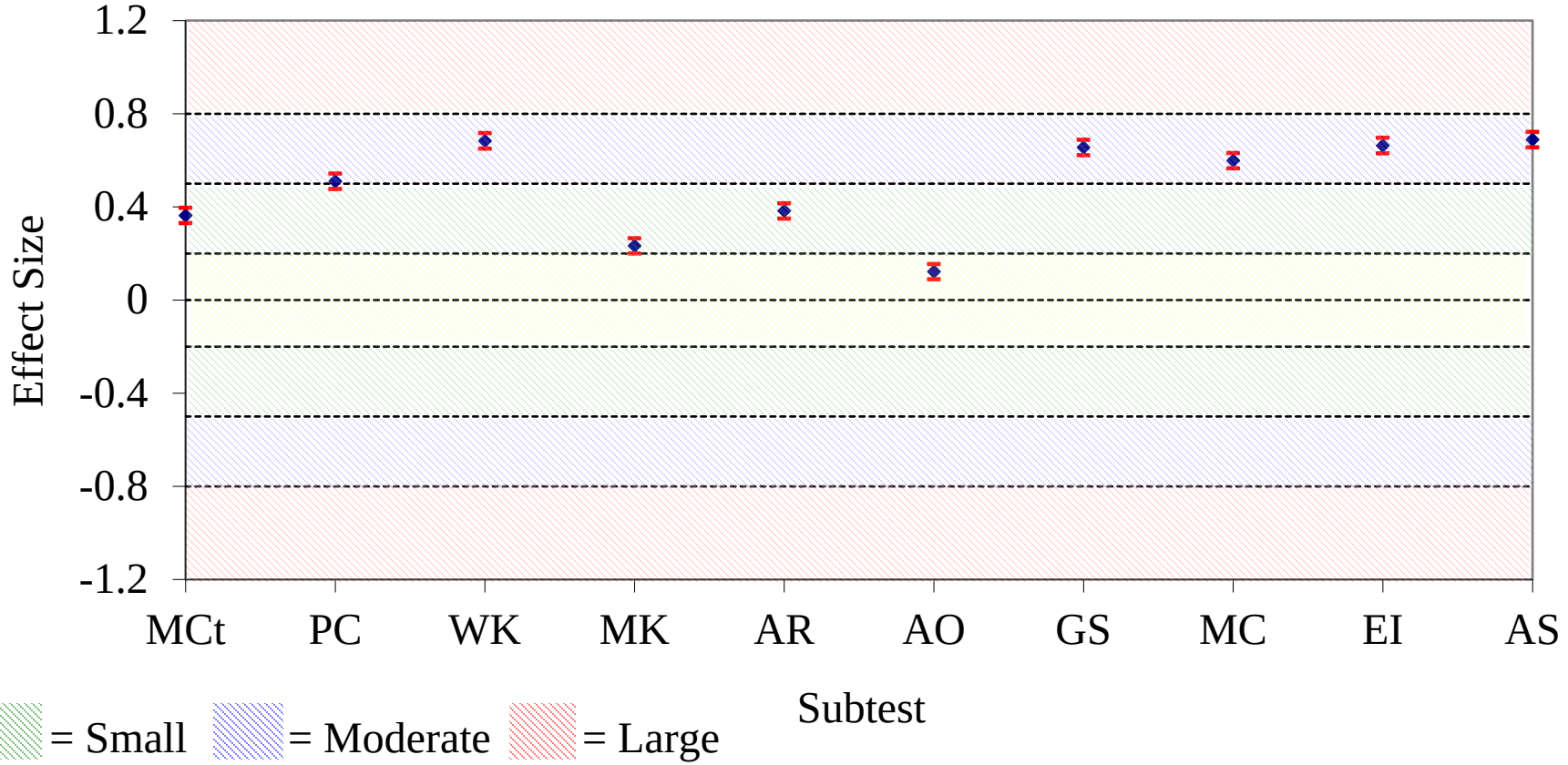
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for Special Tests Scores
Non-Hispanic Whites Versus Hispanic Whites
FY2019



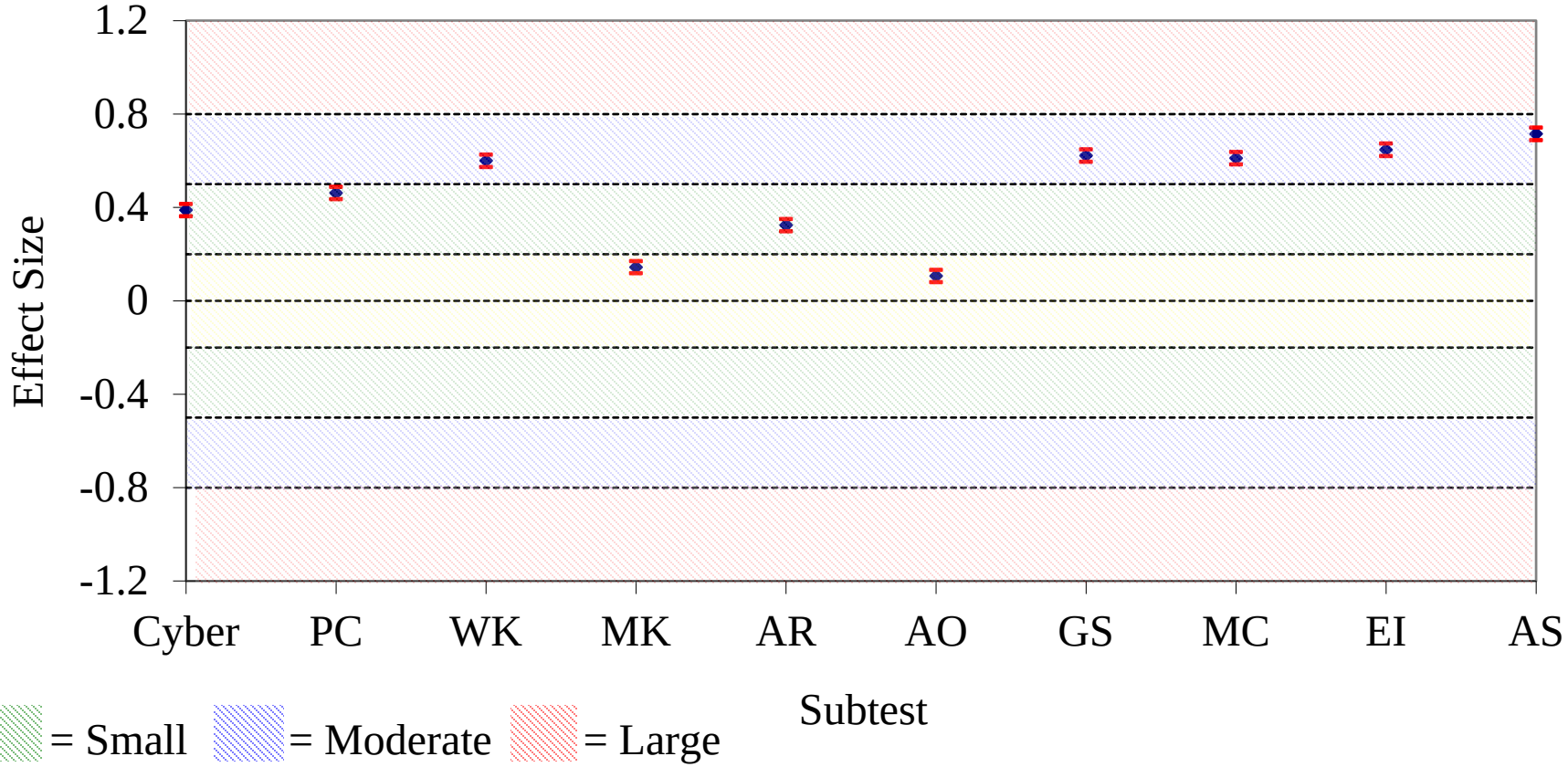
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Hispanic Whites
FY2019, MCt Sample



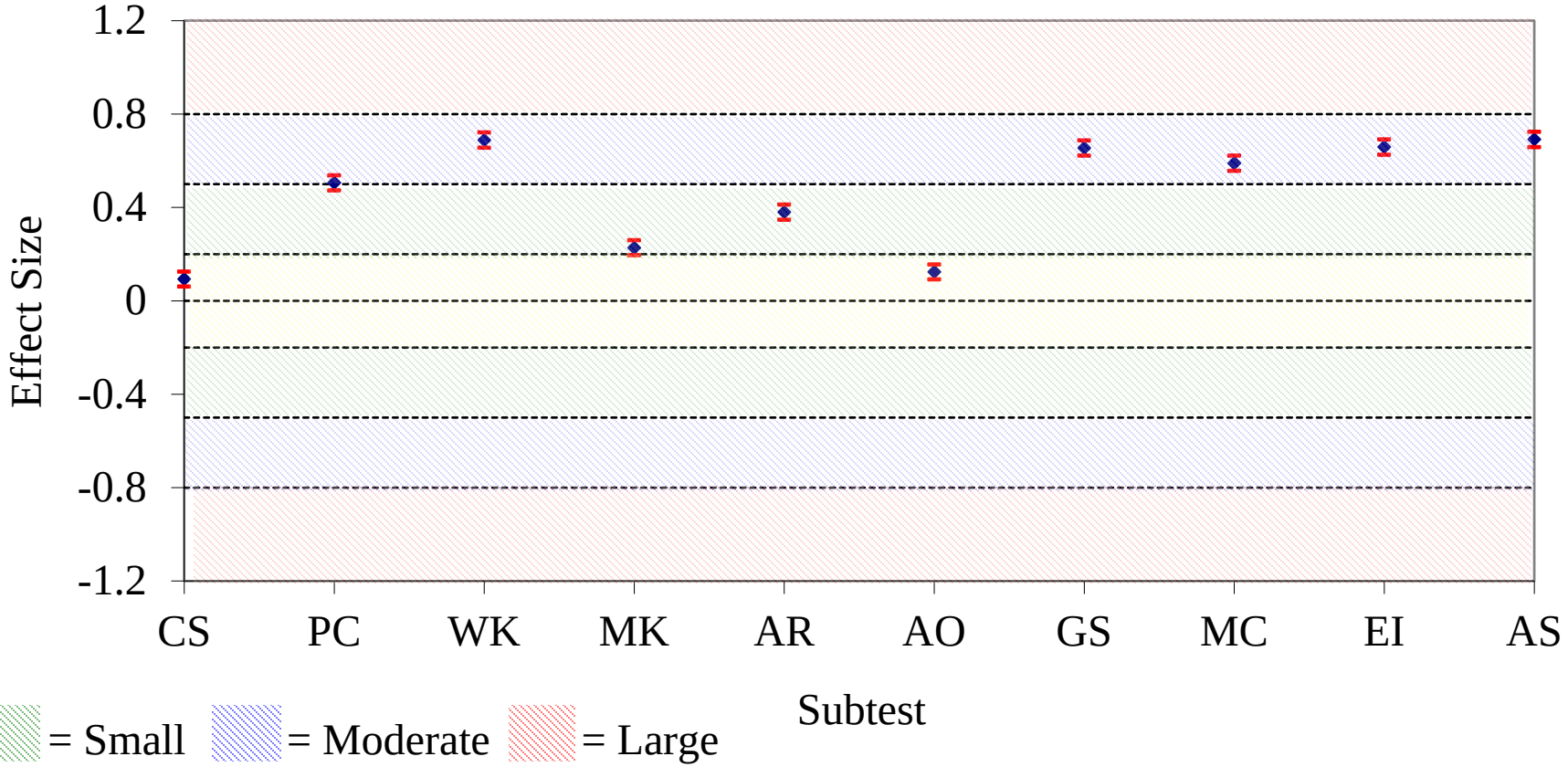
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Hispanic Whites
FY2019, Cyber Sample



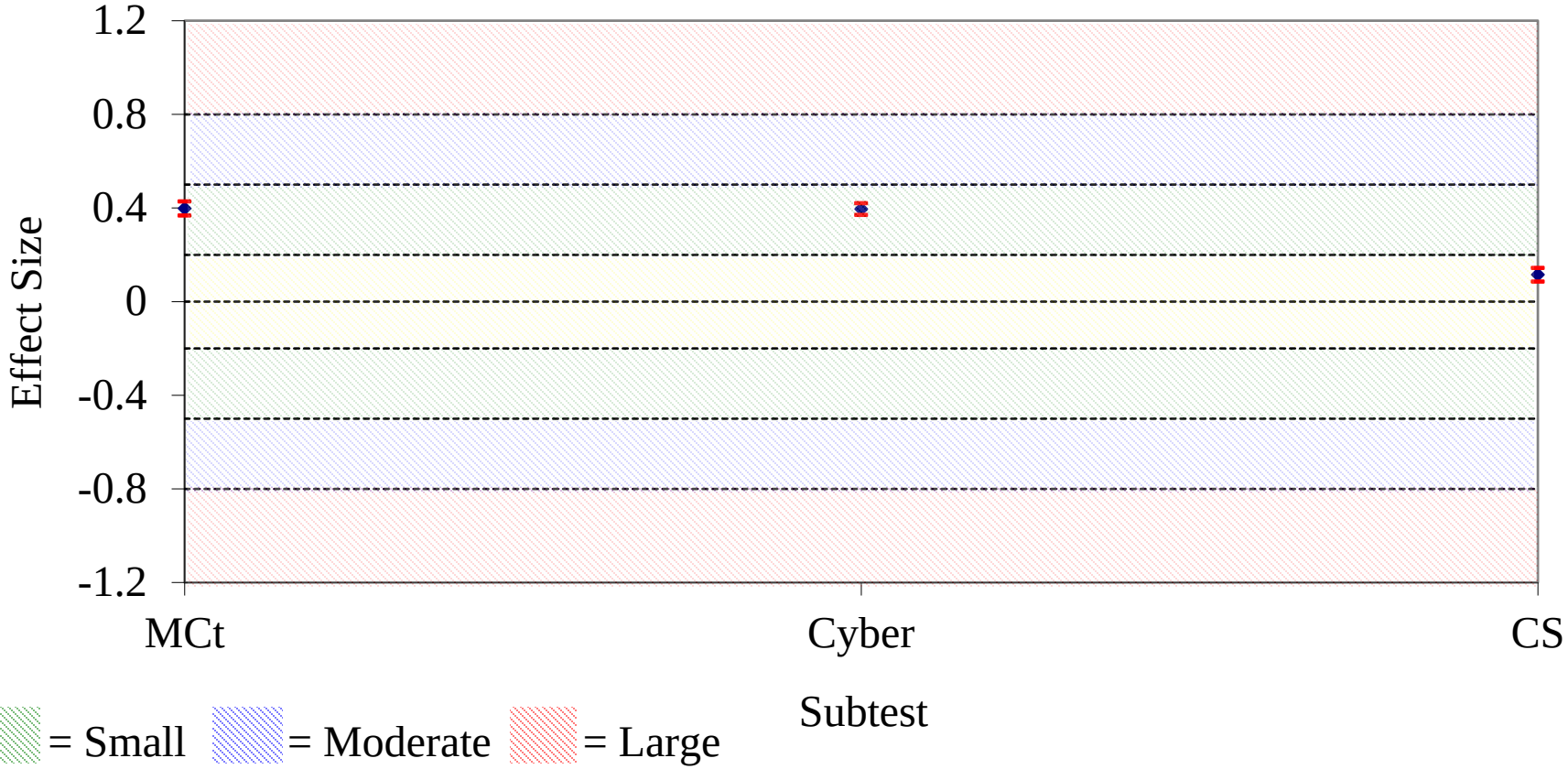
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Hispanic Whites
FY2019, CS Sample



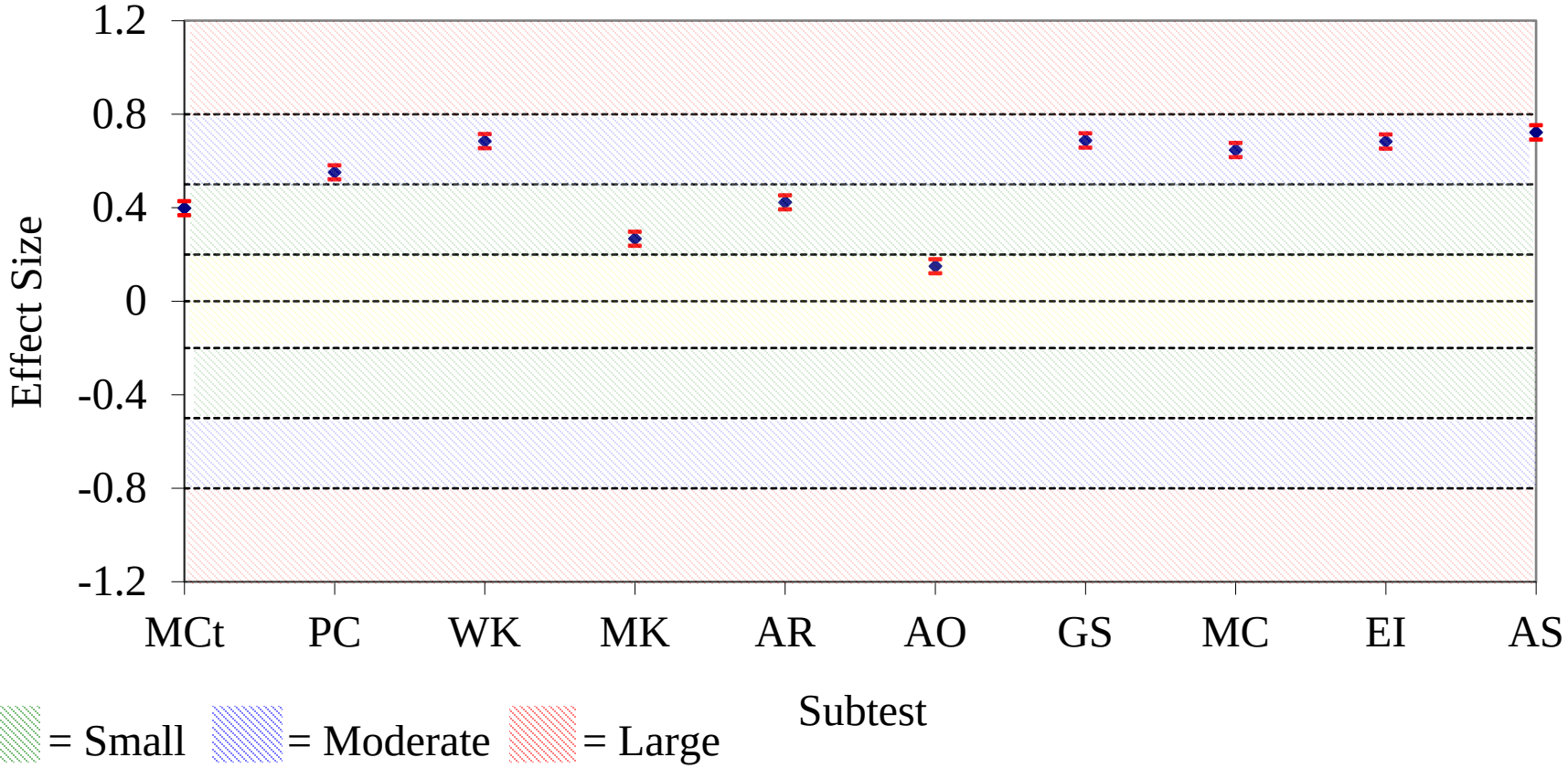
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for Special Tests Scores
Non-Hispanic Whites Versus Hispanics
FY2019



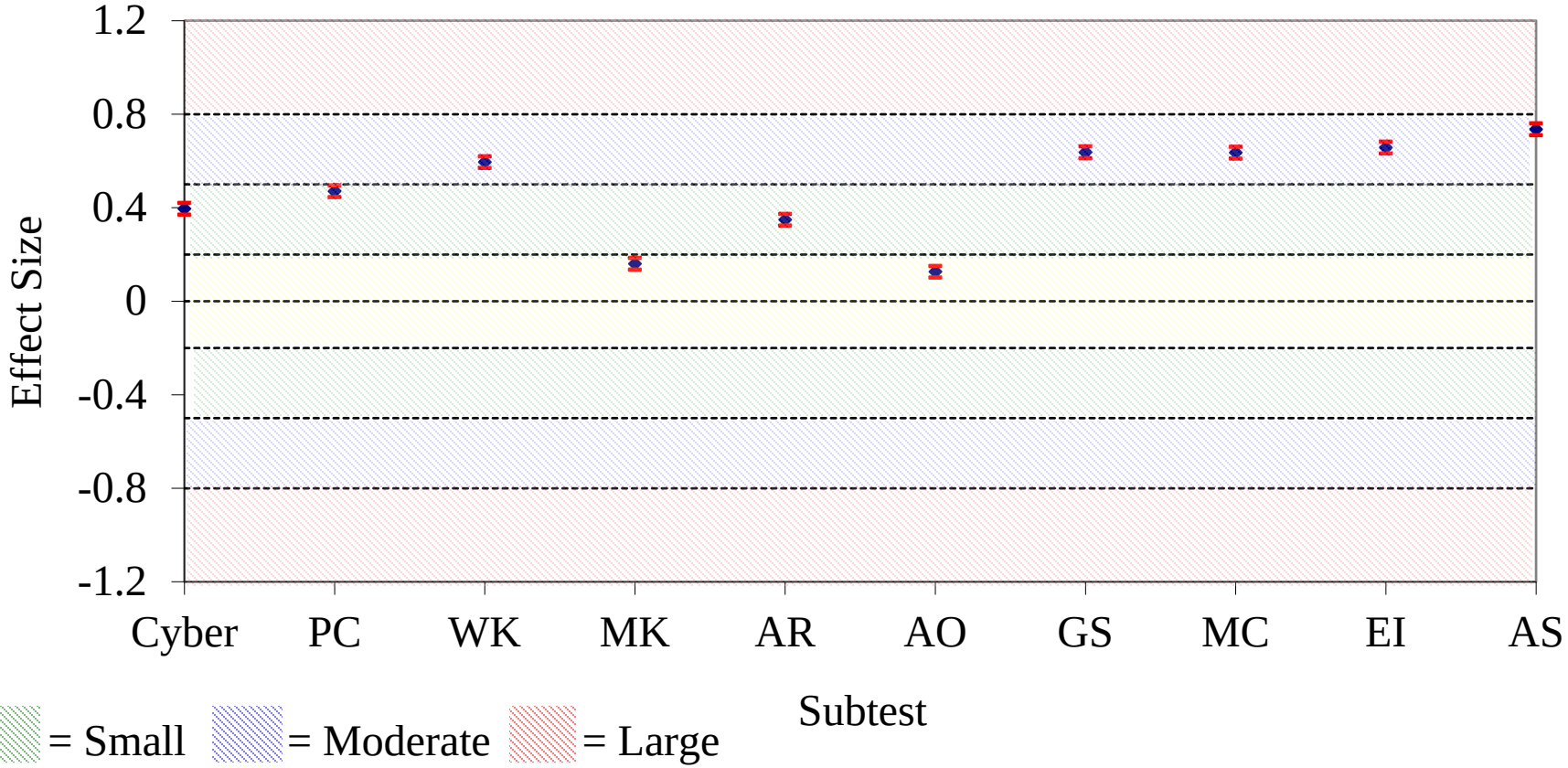
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Hispanics
FY2019, MCt Sample



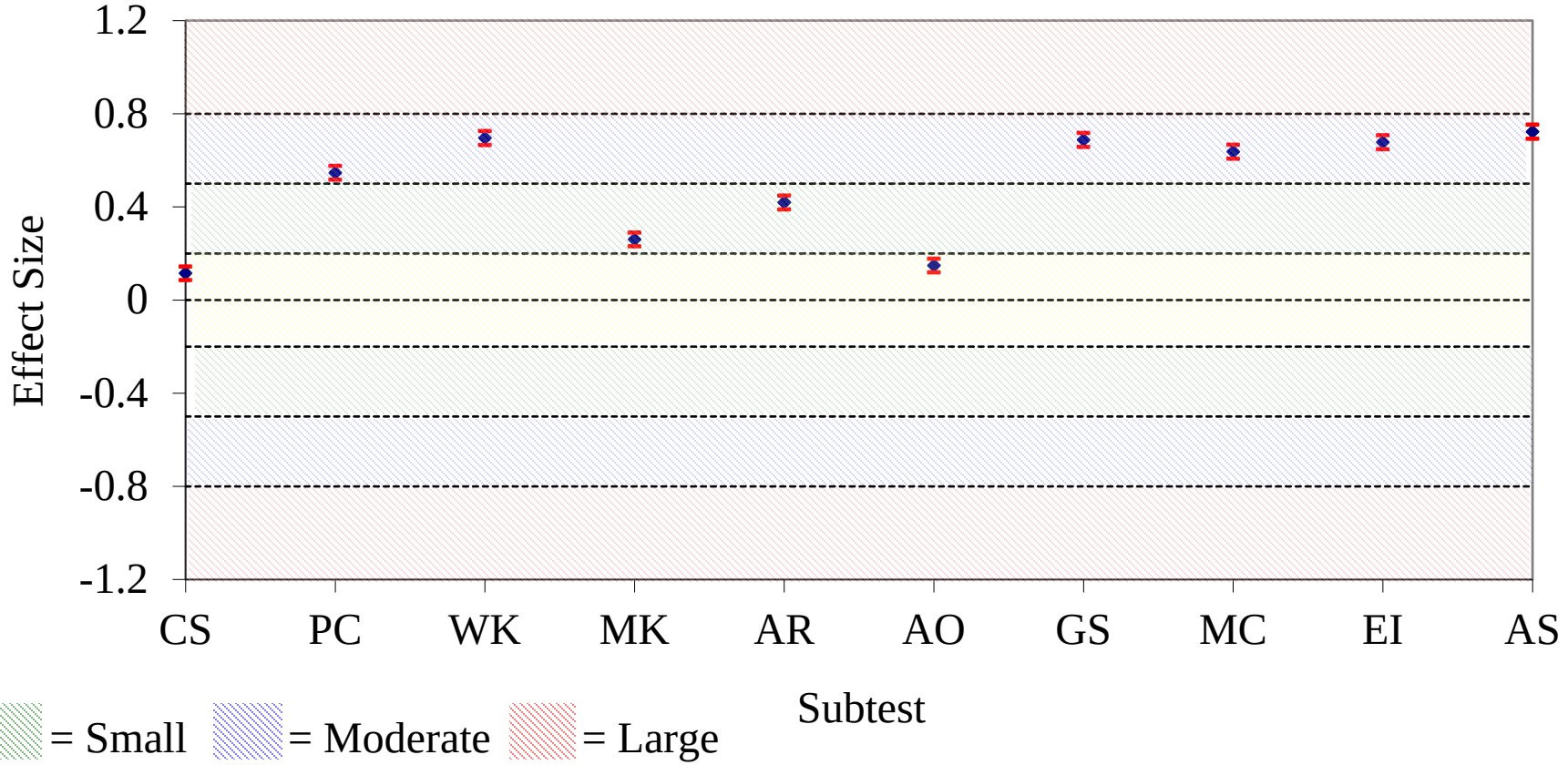
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Hispanics
FY2019, Cyber Sample



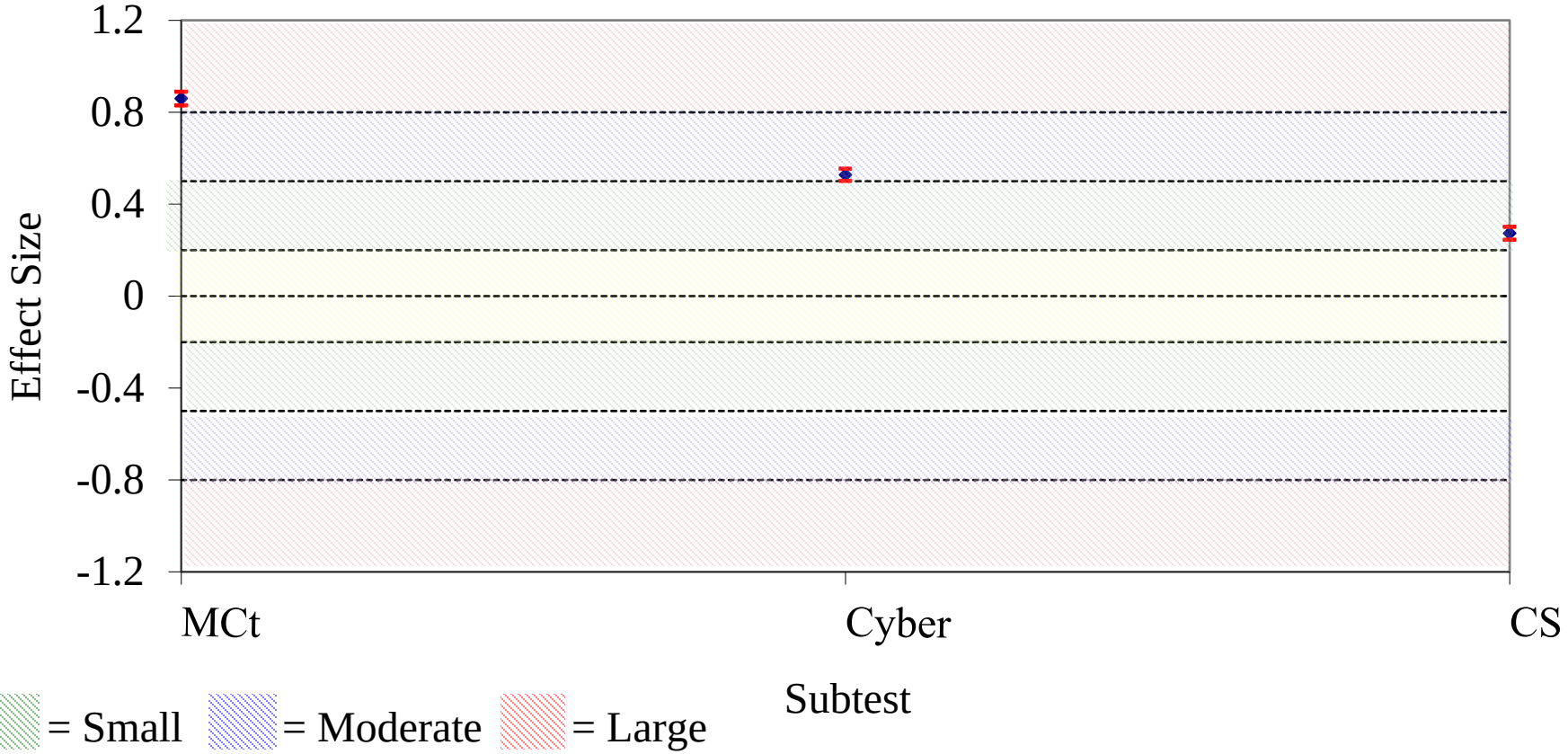
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Hispanics
FY2019, CS Sample



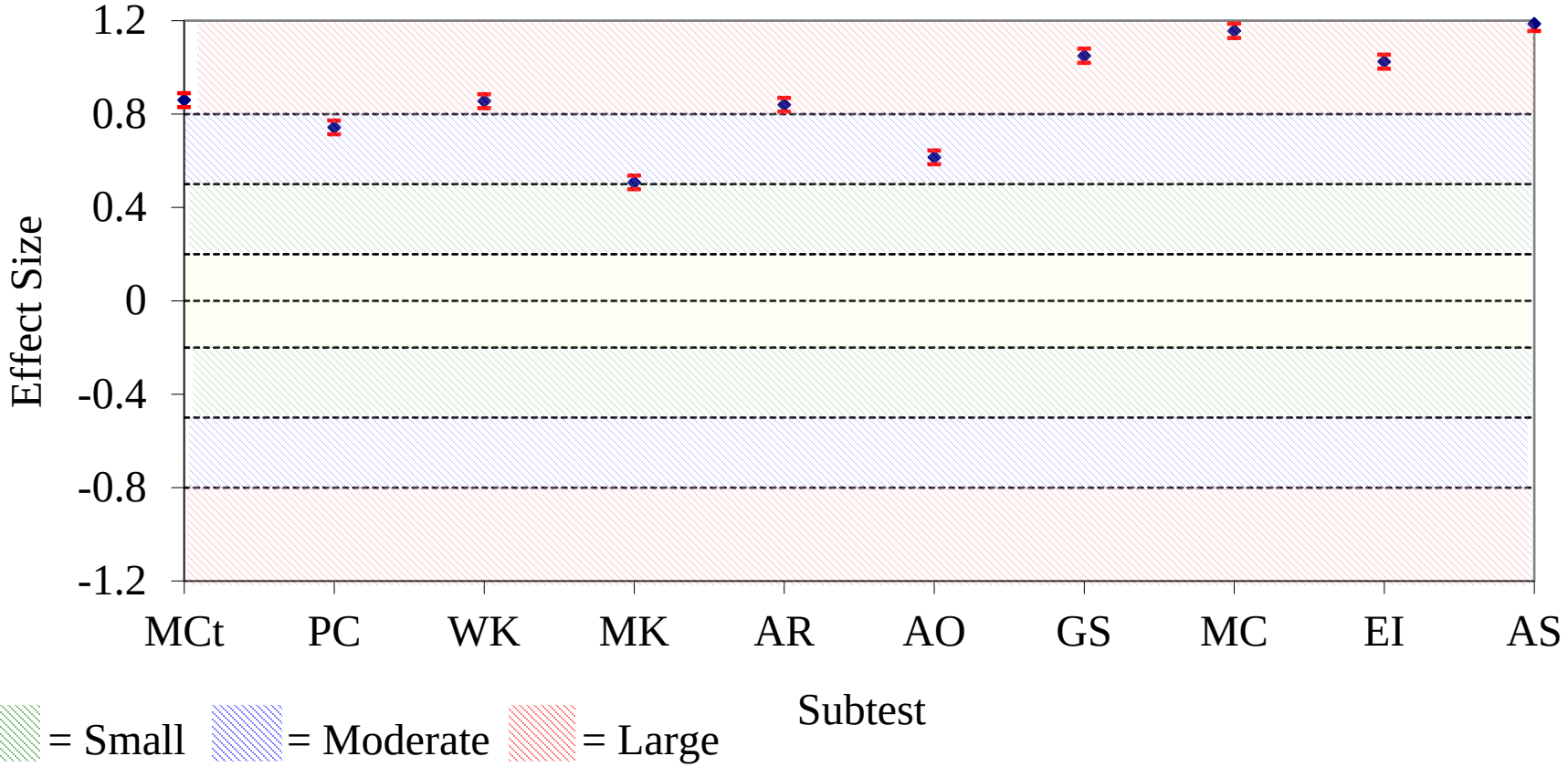
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for Special Tests Scores
Non-Hispanic Whites Versus Non-Hispanic Blacks
FY2019



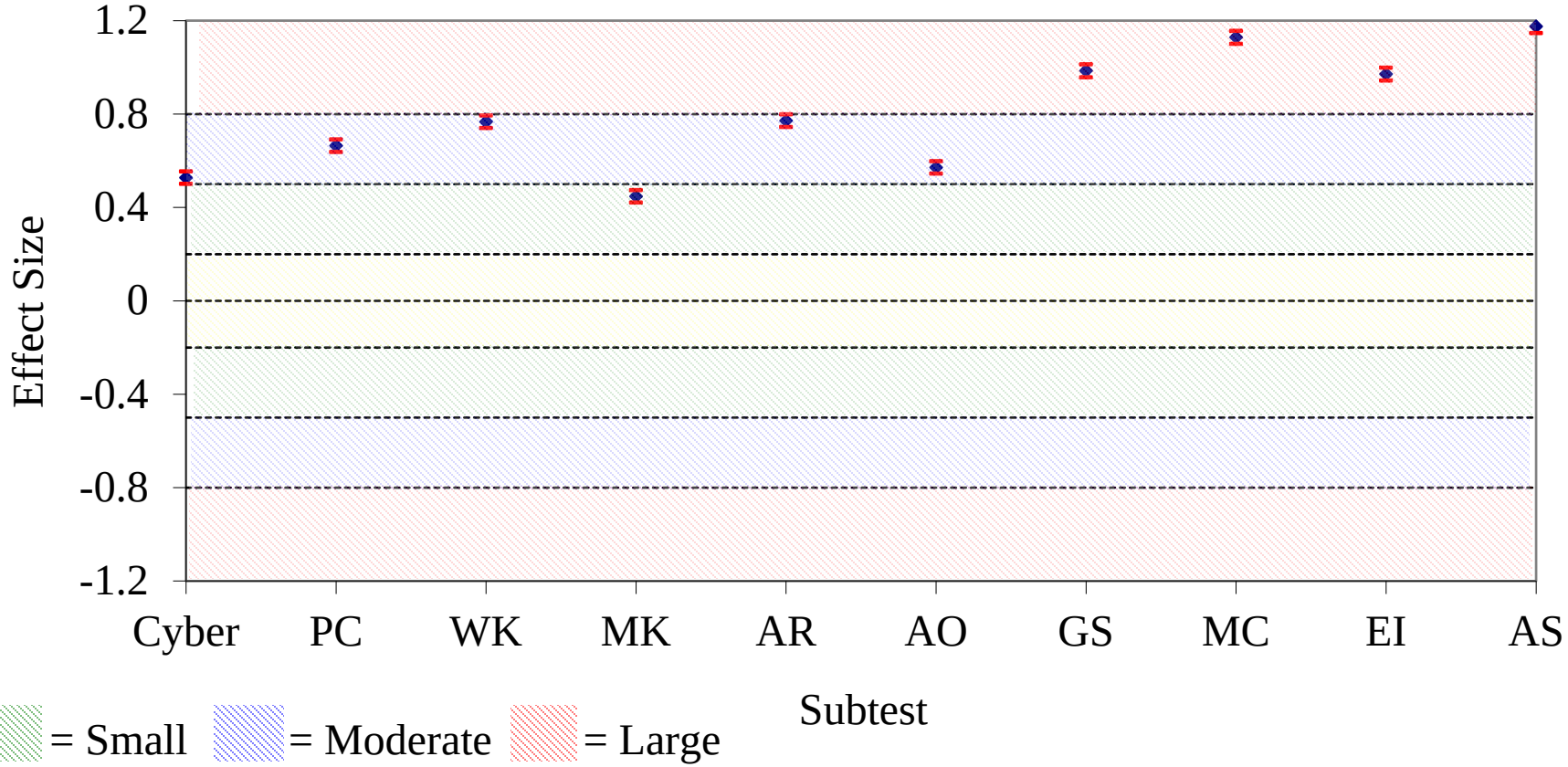
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Non-Hispanic Blacks
FY2019, MCt Sample



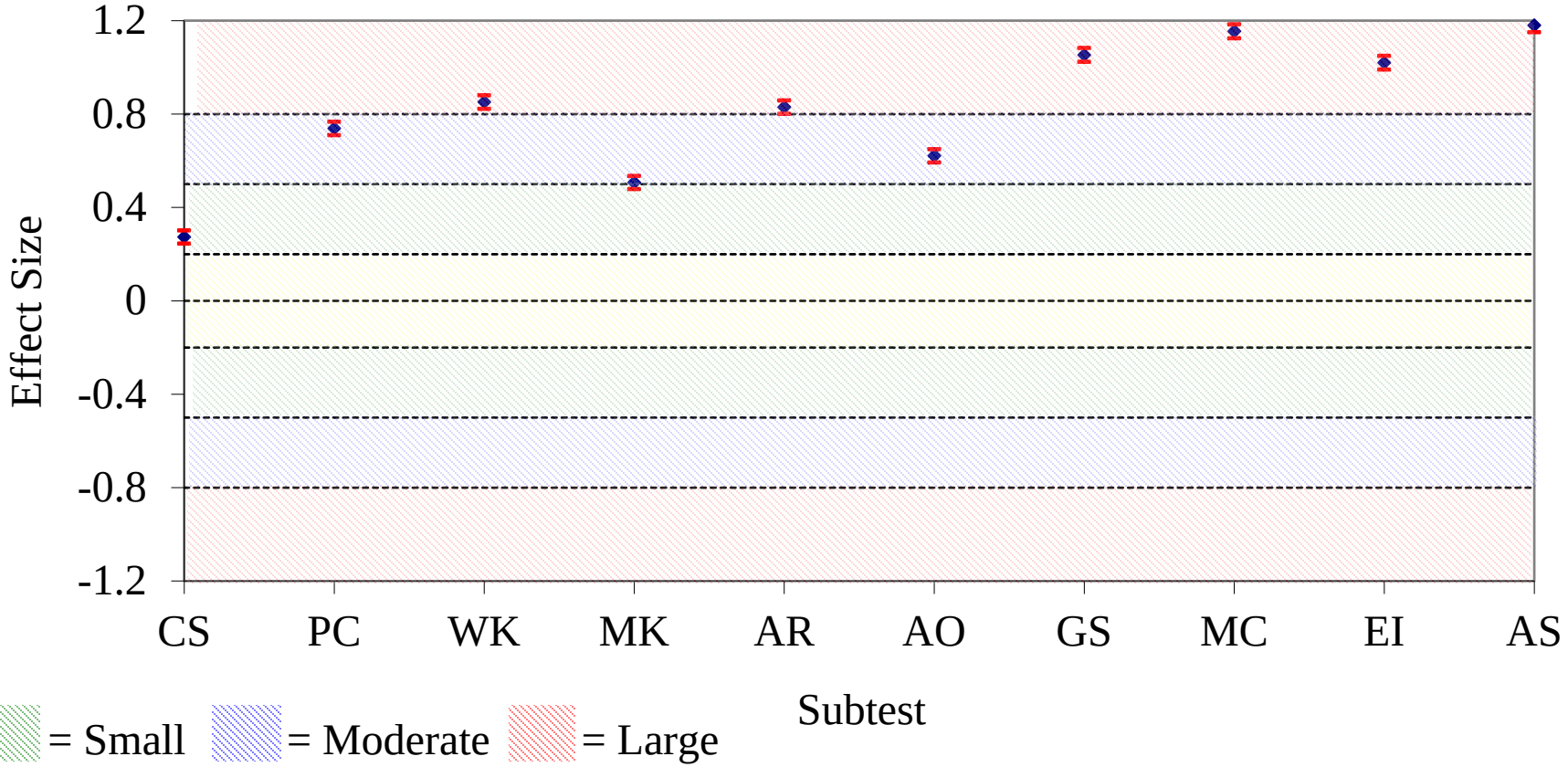
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Non-Hispanic Blacks
FY2019, Cyber Sample



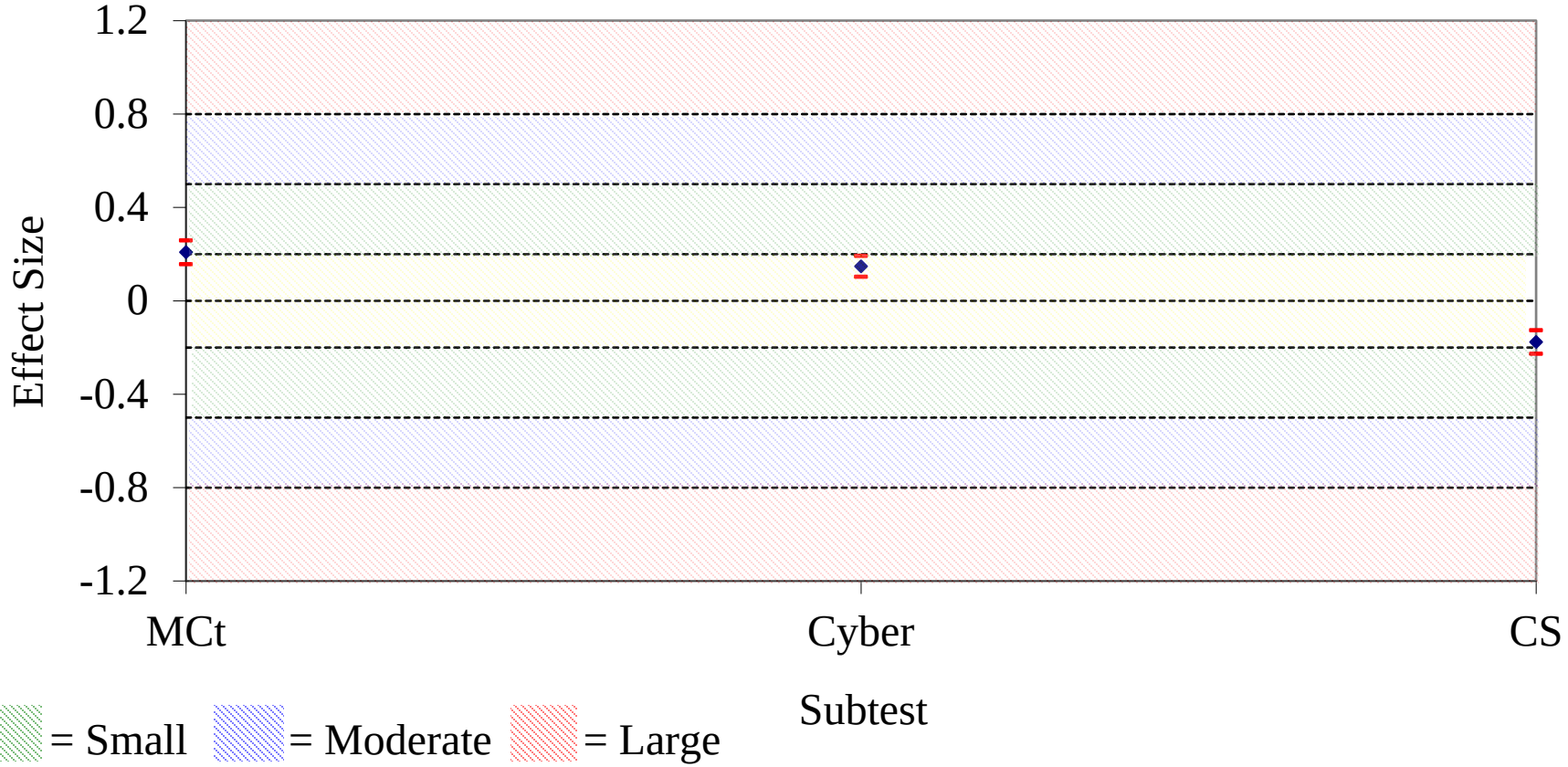
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Non-Hispanic Blacks
FY2019, CS Sample



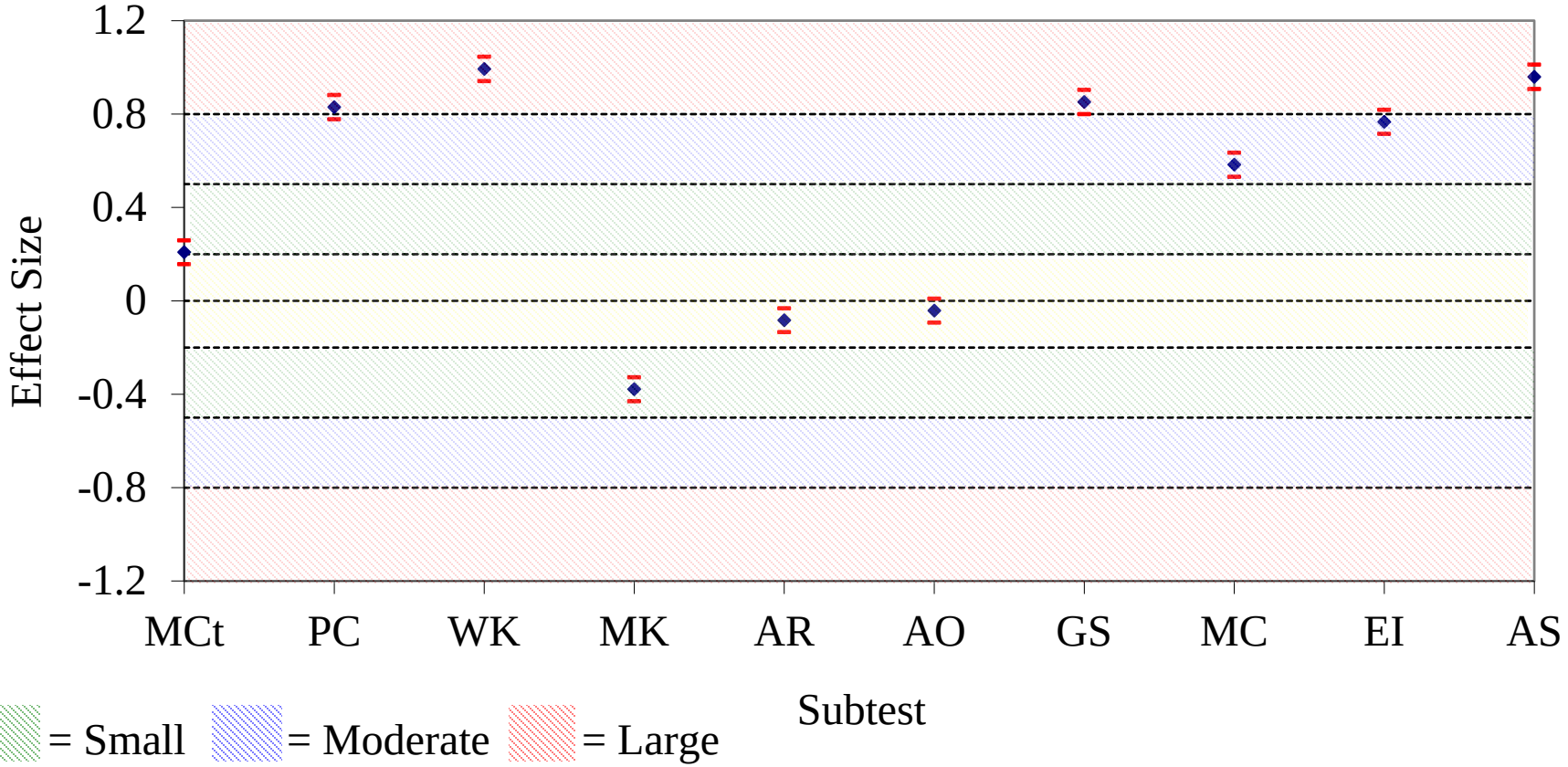
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for Special Tests Scores
Non-Hispanic Whites Versus Non-Hispanic Asians
FY2019



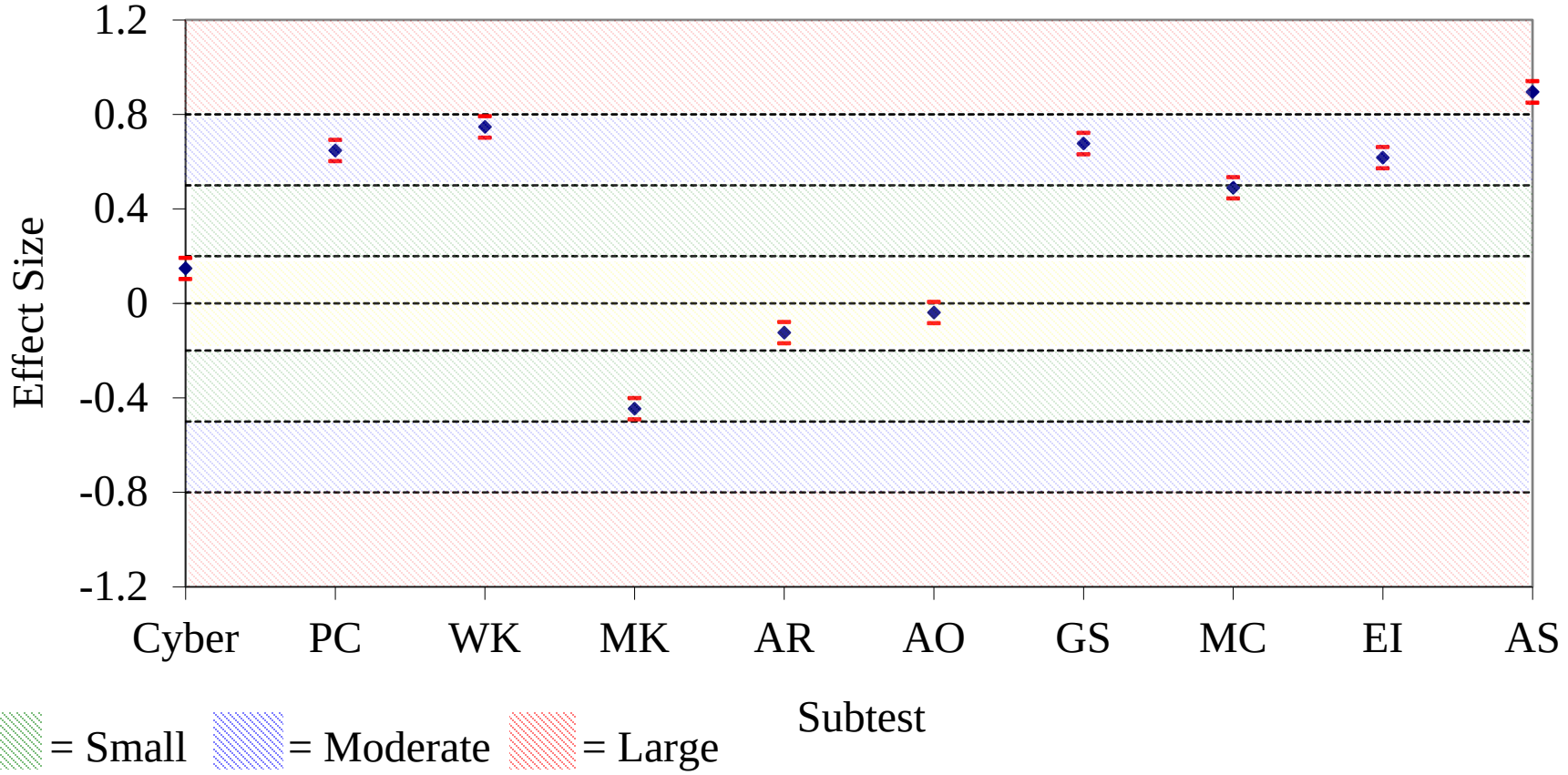
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Non-Hispanic Asians
FY2019, MCt Sample



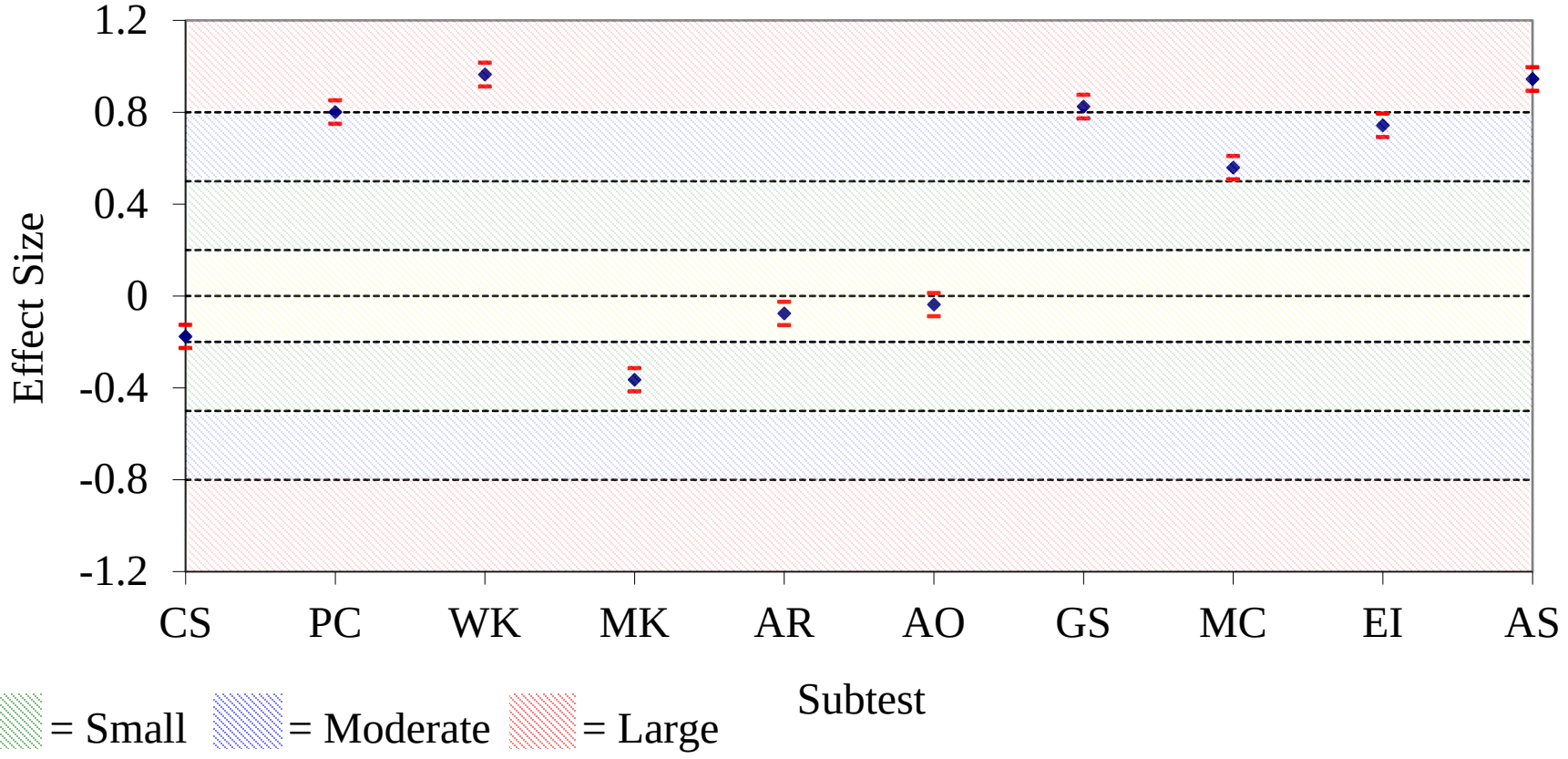
OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Non-Hispanic Asians
FY2019, Cyber Sample



OPA
OFFICE OF PEOPLE ANALYTICS

Effect Sizes (and 95% Confidence Interval) for ASVAB Scores
Non-Hispanic Whites Versus Non-Hispanic Asians
FY2019, CS Sample



OPA
OFFICE OF PEOPLE ANALYTICS

CONCLUSIONS FOR SPECIAL TESTS

- Mental Counters, Cyber Test, and Coding Speed generally exhibited small to moderate effects and were usually as low or lower than most ASVAB tests.
- White-Black comparisons were generally larger for Mental Counters than for the other group comparisons.
- Coding Speed usually had very small effects (near 0), BUT, this test suffers from other issues, for example:
  - Affected by lag time in internet delivery (speeded test)
  - Known to be affected by test delivery device
  - Suffers from coachability and susceptibility to invalid strategies that result in high scores
- Potential for adverse impact is not the only consideration for making changes to the ASVAB.

BACKUP SLIDES

N-SIZE CHART FY2019 ANALYSES

Sample Sizes for FY2019 Analyses

ASVAB sample	N		MCt sample	N		Cyber sample	N		CS sample	N
Males	129,521		Males	23,206		Males	31,673		Males	24,321
Females	43,689		Females	8,573		Females	9,873		Females	8,875
NHW	82,226		NHW	13,685		NHW	19,571		NHW	14,551
HW	32,131		HW	4,933		HW	7,899		HW	5,001
HispanicALL	36,301		HispanicALL	6,362		HispanicALL	9,232		HispanicALL	6,459
NHB	39,915		NHB	7,177		NHB	7,739		NHB	7,487
NHA	7,479		NHA	1,649		NHA	2,121		NHA	1,647

NHW = Non-Hispanic White, HW = Hispanic White, HispanicALL = All Hispanics,
NHB = Non-Hispanic Black, NHA = Non-Hispanic Asian

OPA
OFFICE OF PEOPLE ANALYTICS