OPA

OFFICE OF PEOPLE ANALYTICS

Update on Calculator Impact Study

Kevin Bradley

Human Resources Research Organization

Briefing presented to the DACMPT

January 22, 2025

Briefing Agenda

- Overview
- Results
  - Research Question 1: Does calculator availability meaningfully impact the dimensionality of Arithmetic Reasoning (AR) and Mathematics Knowledge (MK) subtests?
  - Research Question 2: Do psychometric properties differ based on calculator availability?
  - Research Question 3: Does calculator availability impact subgroup performance differences?
  - Research Question 4: Does calculator availability impact the amount of time needed to complete each math subtest?
  - Supplemental Analyses
- General Conclusions and Implications
- Questions for the DAC

OPA
OFFICE OF PEOPLE ANALYTICS

Bottom Line Up Front (BLUF)

- Allowing calculators did not meaningfully impact the underlying dimensionality of the AR and MK subtests. Subtests remained predominantly unidimensional in both the No Calculator and Calculator conditions.
- Providing calculators made some AR items easier, resulting in modest but nonnegligible increases in average AR scores (standardized mean difference = 0.37), but had relatively little effect on MK item difficulty or scores (standardized mean difference = 0.07).
  - Statistical equating will be necessary to maintain statutorily-required AFQT qualification rates and would nullify potential mean score increases.
  - Different degrees of calculator sensitivity across items may create complications for adaptive testing.
- Allowing calculators had no notable impacts on measurement properties such as subtest reliability and item discrimination.
- The impact of calculator availability was generally similar across demographic subgroups.
- Providing calculators reduced the average time spent on AR, but had no impact on time spent on MK.
  - The magnitude of this result was generally similar across demographic subgroups.

Overview

OPA
OFFICE OF PEOPLE ANALYTICS

Overview

- Current ASVAB policy is “no calculators”
- Previous research (Buckland et al., 2021) surveyed subject matter experts (SMEs) across the Services about whether servicemembers are required to apply mathematics knowledge and arithmetic reasoning *without having access* to a calculator or other tool
  - 68% of surveyed military SMEs indicated some form of math, without a calculator, is required in training
  - 56% reported that some form of math, without a calculator, is required on the job
  - Thus, Buckland et al. (2021) recommended the “no calculator” policy continue

Overview (cont.)

- Expressed concerns over current policy with respect to calculators
  - Other national testing programs (e.g., ACT, SAT, GED) allow calculators on the quantitative tests
  - Exclusion of calculators may result in the perception that the ASVAB testing program is not keeping up with trends in assessment
  - High school curricula often allow calculators during instruction and exams
  - Test items requiring manual calculations may result in increased test anxiety as students are not accustomed to performing such calculations without a calculator

Overview (cont.)

- Purpose:
  - Empirically evaluate the impact on examinee test performance and the psychometric properties of the AR and MK subtests when calculators are allowed on the MK and AR subtests of the ASVAB
- Study design considerations:
  - Maximize generalizability to ASVAB applicant population
  - Minimize security risks to existing ASVAB item pools
  - Minimize disruptions to operational testing of applicants
  - Minimize strain or burden on study participants

Overview (cont.)

- Participants were similar to those who take the ASVAB under operational testing conditions, with (relatively) recent operational ASVAB scores
- Designed to be as similar as possible to ASVAB operational testing
- Administered in MEPS by Test Administrators/Test Control Officers
- Included post-test survey (contextual information about participants, motivation, calculator usage)
- Shippers completed the study during a waiting period on their ship day
  - 3,042 participants met all screening criteria (sufficient effort and motivation)
  - 2,870 participants met all screening criteria and were unequivocally matched to their official ASVAB administration
  - Demographic makeup of participant sample is provided in Appendix slides

OPA
OFFICE OF PEOPLE ANALYTICS

Overview (cont.)

- All participants completed the same 30-item AR form and 25-item MK form
- Two conditions: calculator provided/calculator not provided
- To avoid intermingling or “cross-condition” exposure, all participants on a given day assigned to the same condition
  - Odd days (11th, 19th, 25th of month) = calculator not provided
  - Even days (12th, 20th, 30th of month) = calculator provided

OPA
OFFICE OF PEOPLE ANALYTICS

Research Question 1: Does calculator availability meaningfully impact the dimensionality of AR and MK subtests?

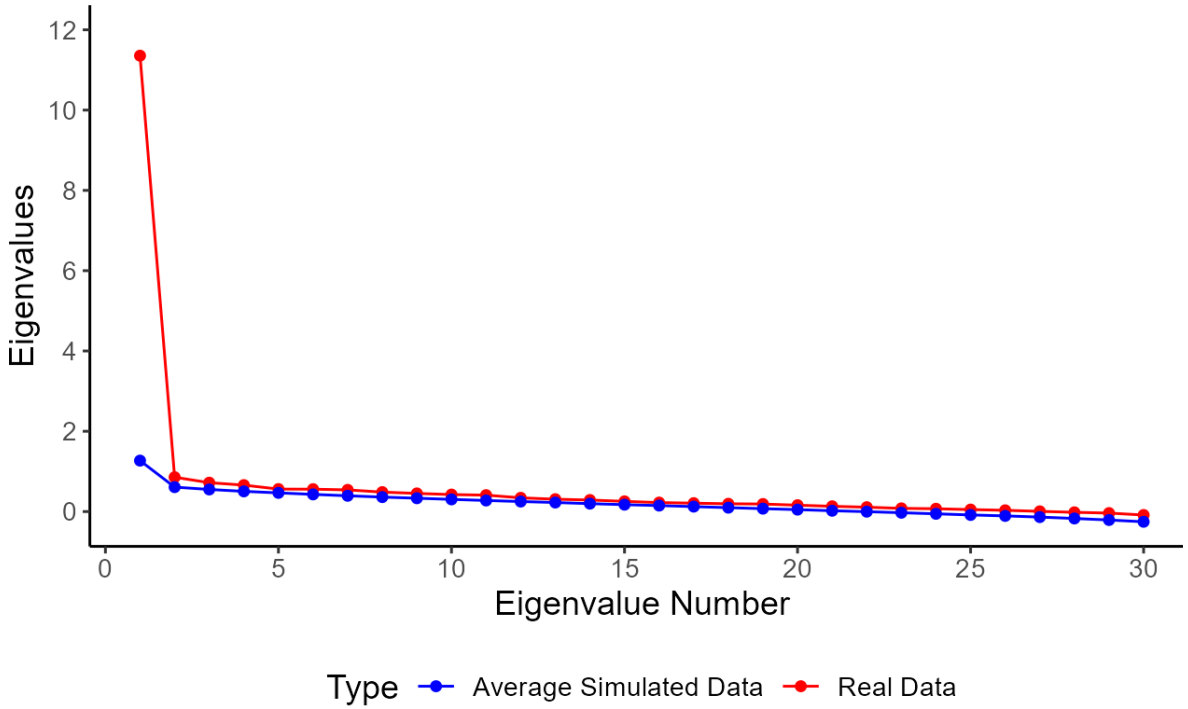
- Parallel analysis
- Bifactor models
- Multiple groups confirmatory factor analysis (CFA)
- Differential functioning of items and tests
- Correlations with other subtest scores

OPA
OFFICE OF PEOPLE ANALYTICS

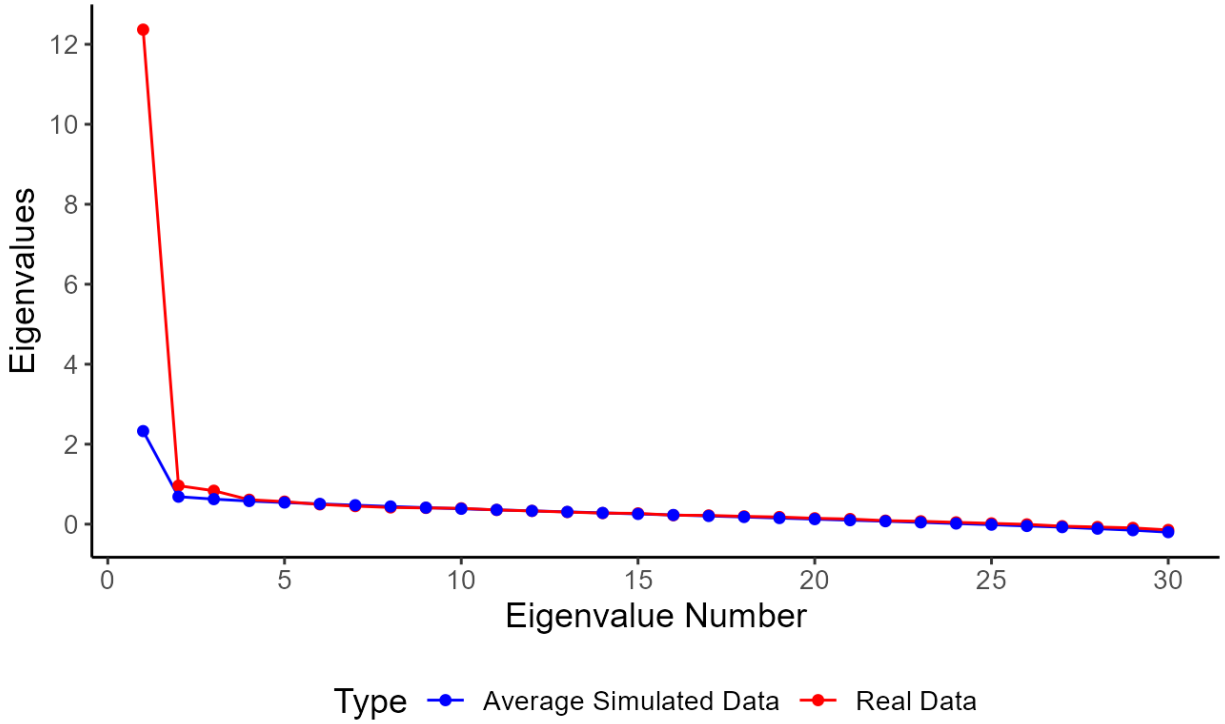
Results

- Parallel analysis results indicate similar AR dimensionality in No Calculator and Calculator conditions

AR No Calculator Condition Scree Plot



AR Calculator Condition Scree Plot

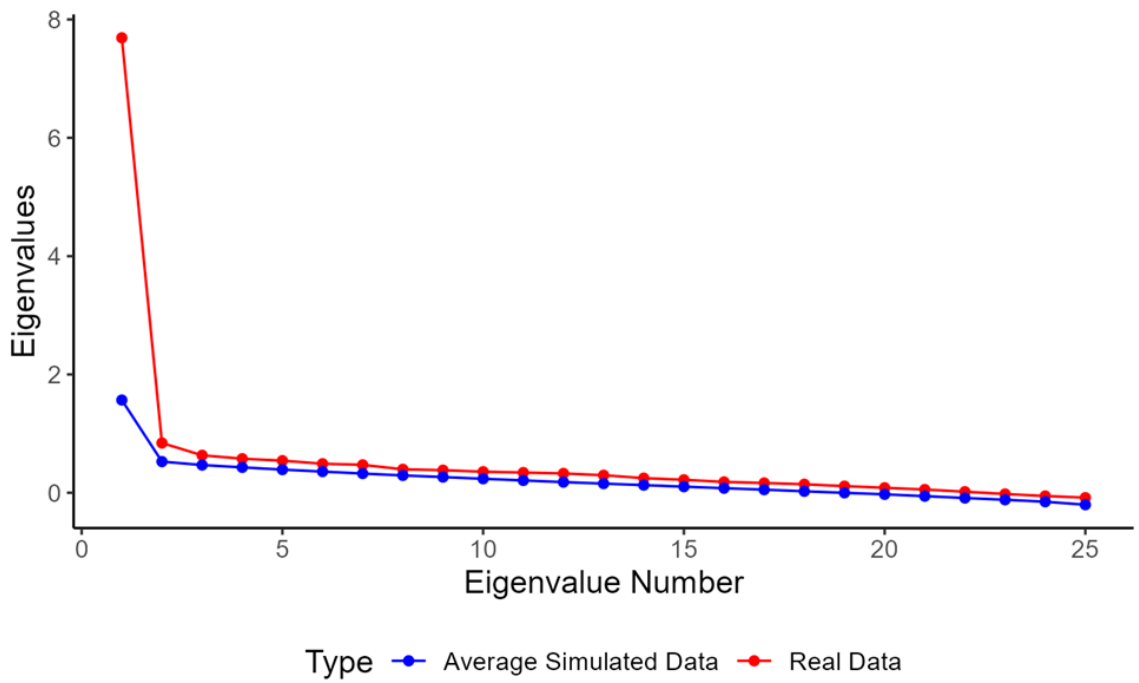


OPA
OFFICE OF PEOPLE ANALYTICS

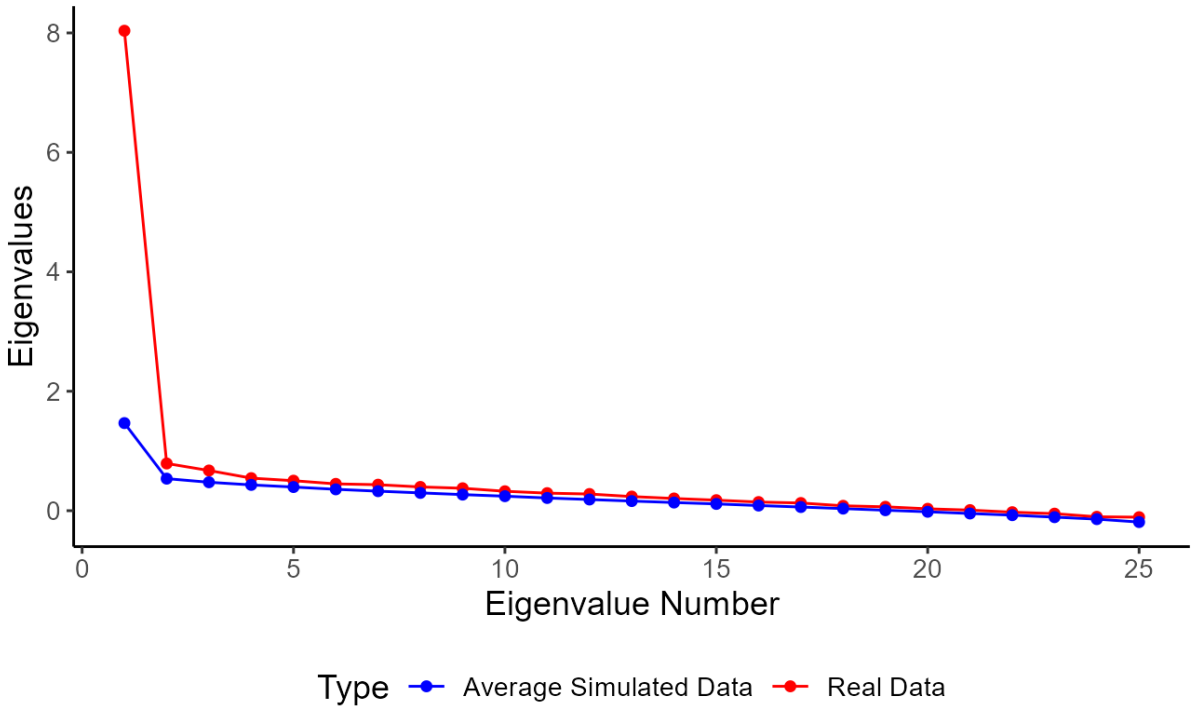
Results (cont.)

- Parallel analysis results indicate similar MK dimensionality in No Calculator and Calculator conditions

MK No Calculator Condition Scree Plot



MK Calculator Condition Scree Plot



OPA
OFFICE OF PEOPLE ANALYTICS

Results (cont.)

- Bifactor model analysis results supported similar dimensionality of AR between conditions, and similar dimensionality of MK between conditions
- Multiple Groups CFA
  - Configural factorial invariance (to test if all items load on a single dimension across groups) was supported for AR and MK
  - Metric (equivalence of factor loadings) invariance partially supported for AR (after removing the equivalence constraints for a subset of items that also demonstrated high non-compensatory differential item functioning [NCDIF] values) and fully supported for MK
  - Scalar (equivalence of intercepts or thresholds) invariance partially supported for AR (after removing the equivalence constraints for a subset of items that also demonstrated high NCDIF values) and MK (after freeing the intercept for one item)

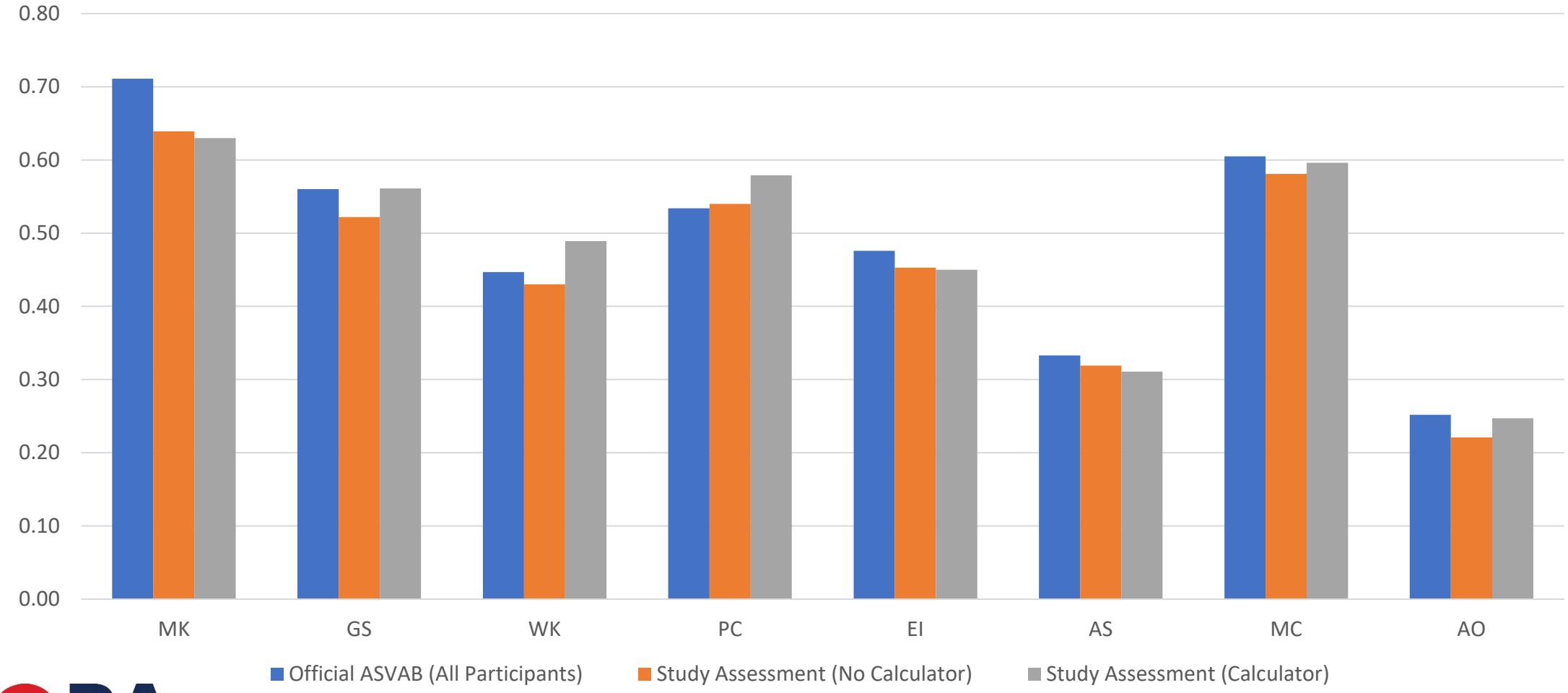
Results (cont.)

- Differential functioning of items and tests (DFIT) test for invariance of item parameters across conditions
  - CDIF & NCDIF
    - 13 AR items and 2 MK items (indicating participants in the Calculator condition more likely to answer correctly)
    - Items exhibiting NCDIF generally exhibited large, positive CDIF values
  - Differential test functioning (DTF)
    - DTF is significant for AR (4.106, compared to significance threshold of .180) but not MK (.068, compared to significance threshold of .150)

OPA
OFFICE OF PEOPLE ANALYTICS

Results (cont.)

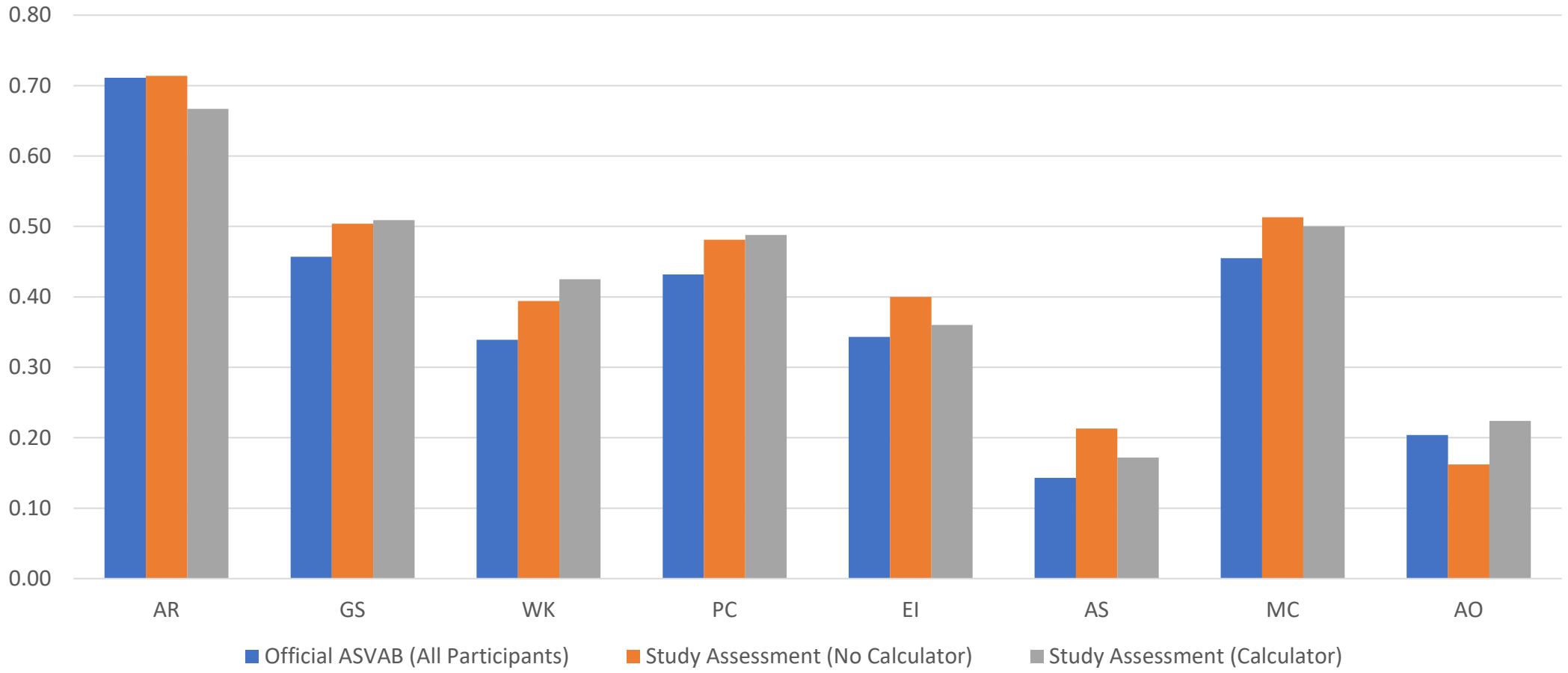
- The pattern and magnitude of AR correlations with other subtest scores are similar for official ASVAB scores, No Calculator condition scores, and Calculator condition scores



OPA
OFFICE OF PEOPLE ANALYTICS

Results (cont.)

- The pattern and magnitude of MK correlations with other subtest scores are similar for official ASVAB scores, No Calculator condition scores, and Calculator condition scores



OPA
OFFICE OF PEOPLE ANALYTICS

Summary and Conclusions for RQ1

- *Does calculator availability meaningfully impact the dimensionality of AR and MK subtests (RQ1)?*
  - Parallel analysis, bifactor models, and CFA results indicate allowing calculators did not meaningfully impact the underlying dimensionality of AR and MK subtests
    - Partial CFA invariance indicates similar factor structure/form across conditions with some items easier in calculator condition
  - DFIT results indicate some items easier in calculator condition
  - Correlations indicate similar patterns across conditions

OPA
OFFICE OF PEOPLE ANALYTICS

Research Question 2: Do psychometric properties differ based on calculator availability?

- Test-level analyses
  - Mean score comparisons
  - Reliability comparisons
  - DTF between conditions
- Item-level analyses
  - Differential item functioning (DIF) between conditions
  - Differences in item statistics between conditions

OPA
OFFICE OF PEOPLE ANALYTICS

Results

- Calculator availability resulted in modest increases in average AR scores but had relatively little effect on MK scores

Subtest	Official Scores					Experimental Scores					Estimated Latent Ability Distributions				
	No Calculator Condition		Calculator Condition		*d*	No Calculator Condition		Calculator Condition		*d*	No Calculator Condition		Calculator Condition		*d*
	M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	
AR	52.74	8.10	52.26	8.09	−0.06	48.88	9.13	52.26	9.23	0.37	48.55	10.34	52.67	10.56	0.39
MK	53.21	7.06	52.97	6.81	−0.03	49.20	7.04	49.66	7.09	0.07	49.11	7.60	49.64	7.67	0.07

OPA
OFFICE OF PEOPLE ANALYTICS

Results (cont.)

- Allowing calculators had no notable impact on subtest reliability

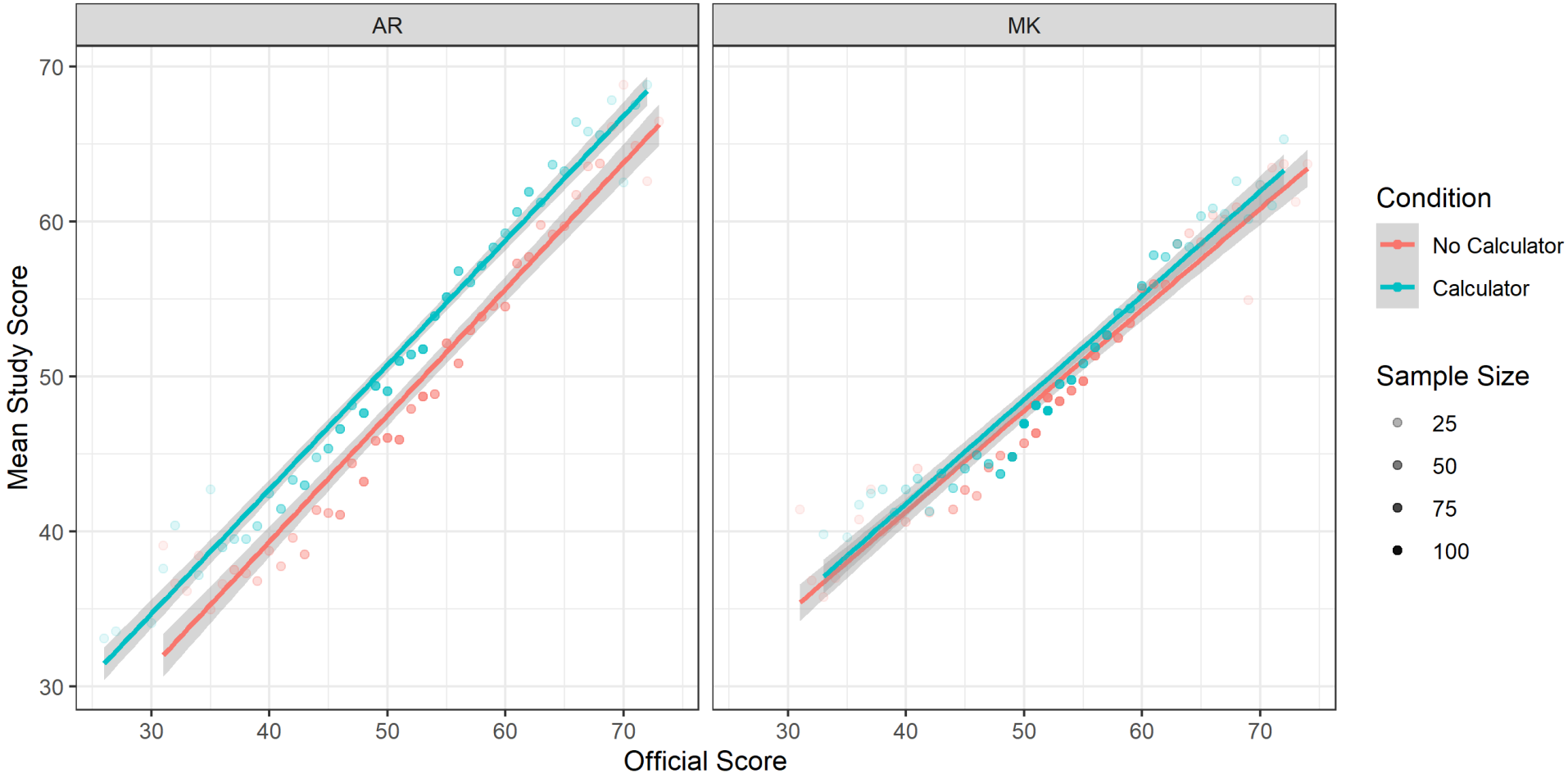
Reliability Type	Method	Subtest	Condition		
			No Calculator Condition	Calculator Condition	Difference
Coefficient alpha	Unstandardized	AR	.895	.902	.007
		MK	.846	.852	.006
	Standardized	AR	.894	.901	.008
		MK	.845	.852	.006
Marginal IRT R_{xx}	Empirical	AR	.858	.864	.006
		MK	.791	.799	.008
	Projected	AR	.873	.868	−.005
		MK	.847	.865	.018

Note: Empirical marginal IRT reliability are based on participants' theta estimates and corresponding standard errors; projected IRT reliability used estimated IRT parameters (no calculator scale) and N(0,1) ability distribution.

OPA
OFFICE OF PEOPLE ANALYTICS

Results (cont.)

Relations between official scores and study scores by condition (DTF between conditions)



Note: For AR, the results indicate that mean study scores (conditional on official scores) are higher for Calculator condition participants compared to No Calculator condition participants. The AR results indicate a simple main effect of calculators on participants' scores. No such pattern was observed for MK, as the conditions' regression lines were not statistically significantly different.

OPA
OFFICE OF PEOPLE ANALYTICS

Results (cont.)

- AR items were generally easier for participants in the Calculator condition (b parameter estimates on the No Calculator scale)
  - Linear equating nullified the mean difficulty differences between conditions
  - Items' a and c parameters were not systematically different between conditions
- MK items were not systematically easier when a calculator was available

3PL Item Parameter	Subtest	No-Calculator Scale				
		No Calculator Condition		Calculator Condition		*d*
		M	*SD*	*M*	*SD*	
a	AR	1.35	0.39	1.43	0.46	0.18
	MK	1.35	0.40	1.46	0.45	0.27
b	AR	0.02	0.47	-0.31	0.53	-0.67
	MK	0.56	0.32	0.53	0.39	-0.10
c	AR	0.22	0.07	0.23	0.06	0.12
	MK	0.22	0.08	0.23	0.10	0.11

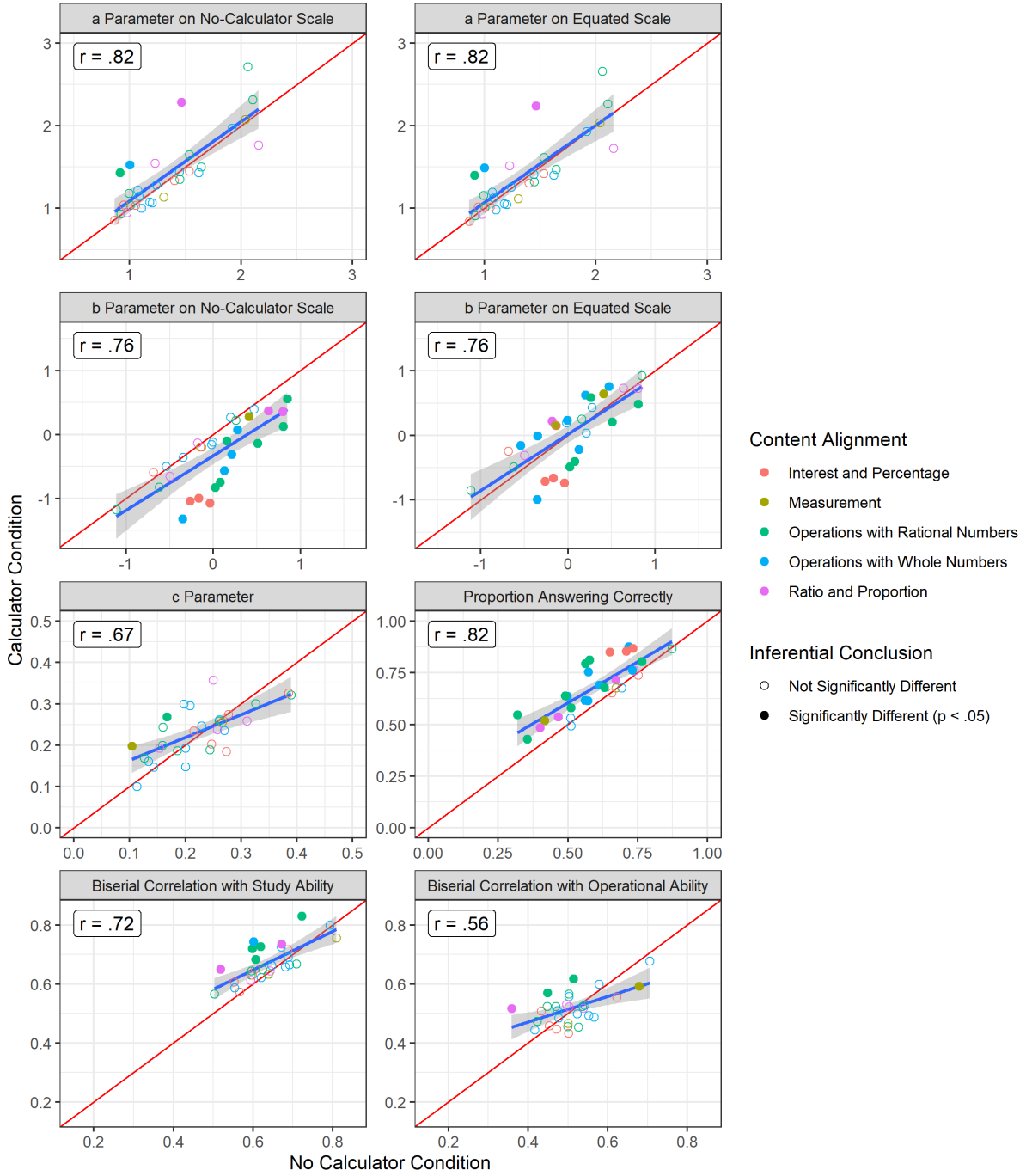
3PL Item Parameter	Subtest	Equated Scale				
		No Calculator Condition		Calculator Condition		*d*
		M	*SD*	*M*	*SD*	
a	AR	1.35	0.39	1.40	0.45	0.11
	MK	1.35	0.40	1.45	0.45	0.24
b	AR	0.02	0.47	0.04	0.54	0.03
	MK	0.56	0.32	0.58	0.39	0.06
c*	AR	0.22	0.07	0.23	0.06	0.12
	MK	0.22	0.08	0.23	0.10	0.11

*c parameter is not transformed

Results (cont.)

- Relations between item-level statistics from study conditions for AR

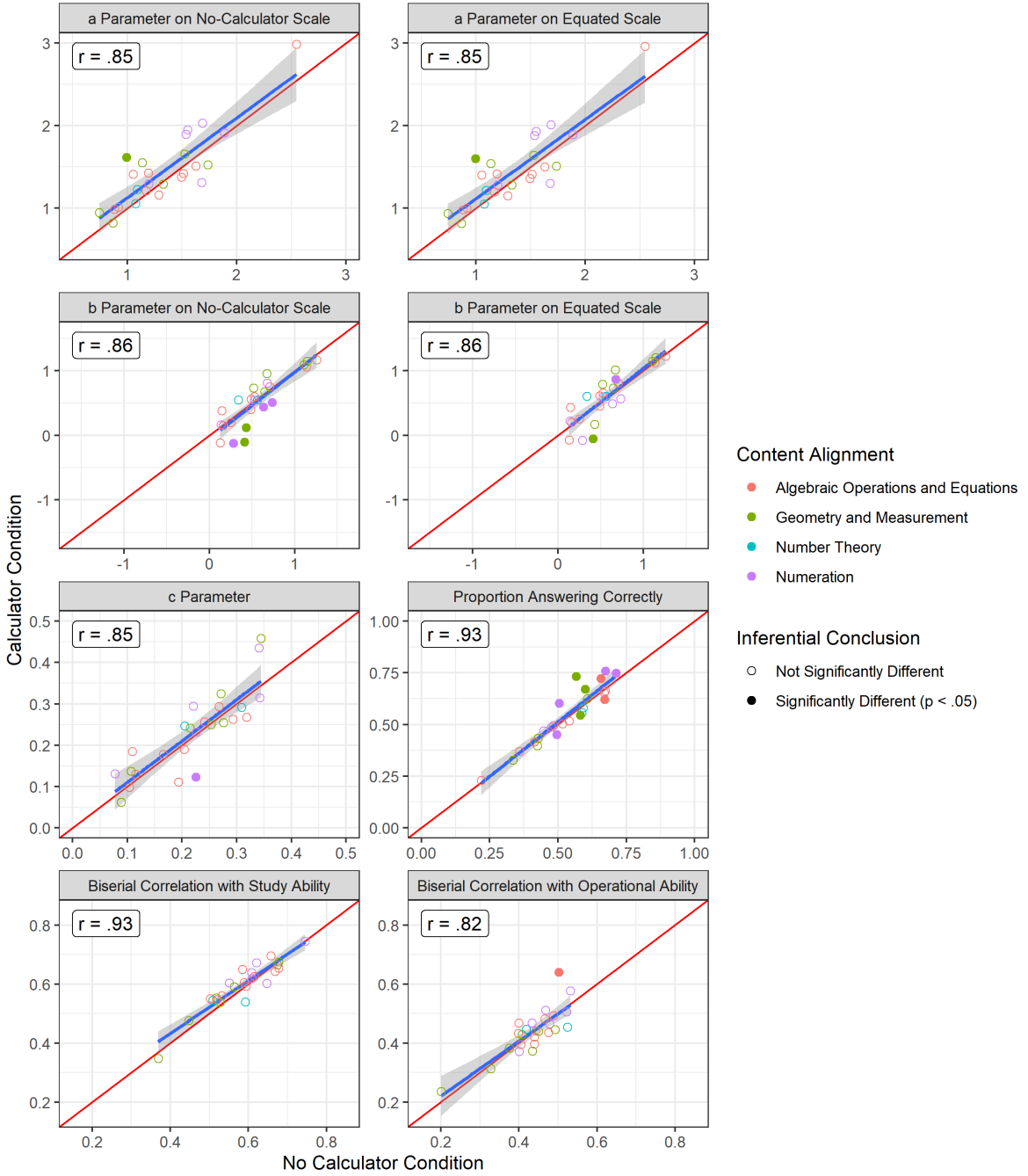
Note: Difficulty parameters and p-values reflect some items were easier in the Calculator condition. All plots reflect a strong positive relationship between item statistics in the No Calculator and Calculator conditions.



Results (cont.)

- Relations between item-level statistics from study conditions for MK

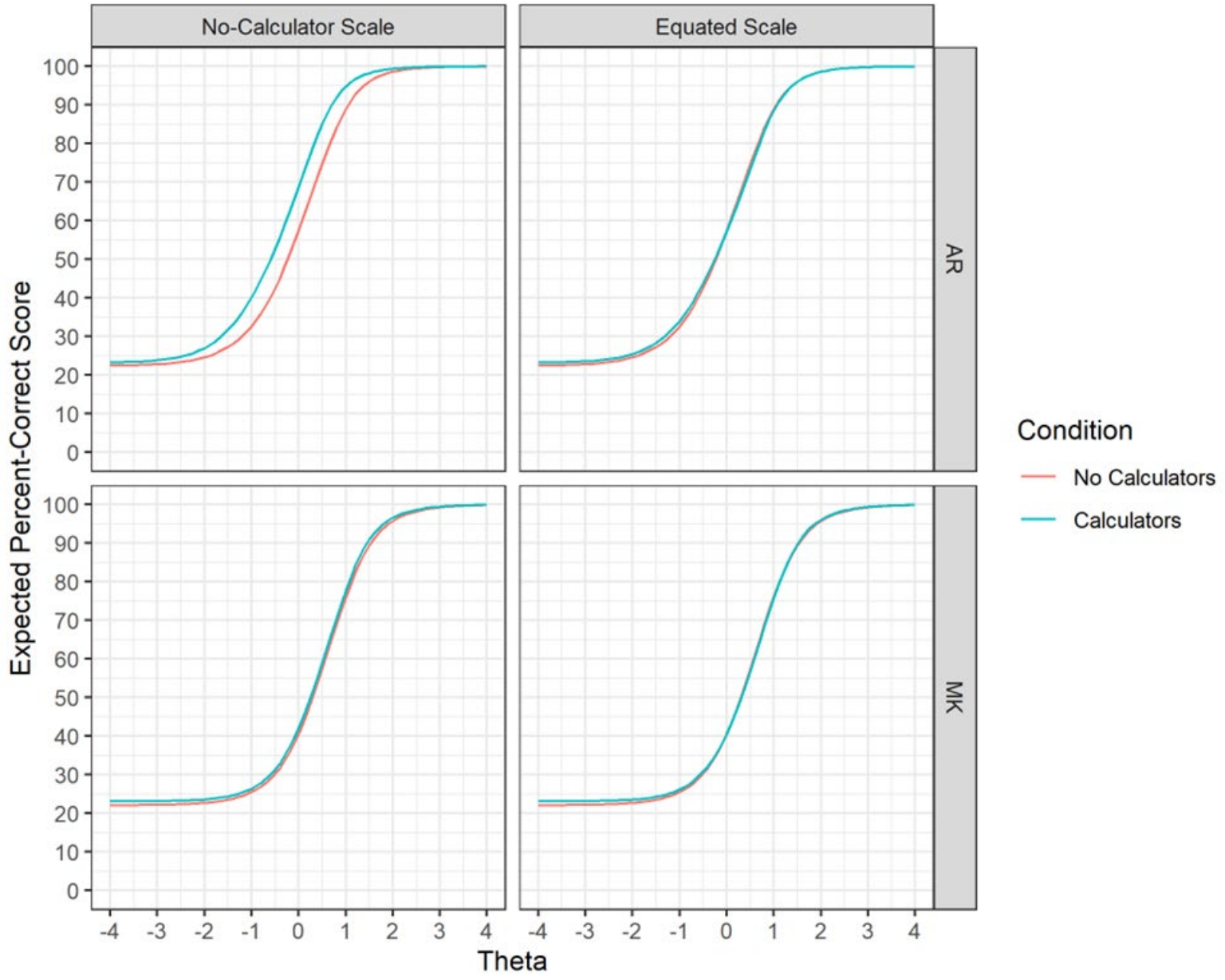
Note: All plots reflect a strong positive relationship between item statistics in the No Calculator and Calculator conditions.



Results (cont.)

- IRT test characteristic curve (TCC) comparisons between conditions

Note: Differences in AR TCCs between conditions were minimal after using linear rescaling to account for the impact of calculators on estimated latent ability distributions.



OPA
OFFICE OF PEOPLE ANALYTICS

Summary & Conclusions for RQ2

- *Do psychometric properties differ based on calculator availability (RQ2)?*
  - Calculators make some AR items easier, but have very little impact on the difficulty of MK items
  - The effects of calculators on scores and item difficulty parameters are primarily linear (after equating, TCCs for No Calculator and Calculator conditions are nearly identical)
    - No Calculator and Calculator conditions could be linked through linear rescaling procedures applied to either scores or item parameters to maintain the interpretability of standard scores and composite scores
      - This finding is likely limited to the individually equated, fixed linear forms used in this study (we do not expect it to generalize to all P&P-ASVAB forms, nor to CAT-ASVAB forms)
      - Even though the mean effects of calculators on item parameters were nullified via IRT equating, there was considerable variance in the differences in AR items' equated b parameters between conditions, and a few items had outlier a parameters in the Calculator condition
      - There was less variance in MK items' parameter differences between conditions, but our DIF analyses showed that a small proportion of MK items are likely to be calculator sensitive
      - This variance in equated item parameters means that a CAT assessment based on equated parameters might encounter inefficiencies due to items' actual parameters differing from the equated parameter estimates
    - Equating would be an essential component of introducing calculators to operational ASVAB testing (to maintain continuity of scores), **resulting in no systematic advantage gained by examinees from using calculators**

Research Question 3: Does calculator availability impact subgroup performance differences?

- Mean score differences across subgroups
- Adverse impact potential by condition
- Within condition DIF analysis

OPA
OFFICE OF PEOPLE ANALYTICS

Results

- The magnitudes of effect sizes between conditions are consistent across subgroups

Subgroup	No Calculator Scores					Calculator Scores					Effect Size	
	n	AR		MK		*n*	AR		MK		AR	MK
		M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	*d*	*d*
Overall	1,382	48.88	9.13	49.20	7.04	1,488	52.26	9.23	49.66	7.09	.37	.07
Female	158	45.86	7.76	47.89	6.42	169	48.98	9.11	48.80	6.76	.37	.14
Male	1,216	49.23	9.19	49.34	7.11	1,307	52.60	9.21	49.77	7.14	.37	.06
Hispanic White	291	47.65	8.34	47.85	6.87	363	51.18	8.30	48.57	6.78	.43	.11
Non-Hispanic Asian	50	51.30	9.10	52.21	7.32	68	52.89	9.77	51.60	8.77	.17	−.07
Non-Hispanic Black	326	44.81	8.17	47.30	6.26	313	48.34	8.39	47.78	6.55	.43	.07
Non-Hispanic White	629	51.54	9.07	50.78	6.98	652	54.71	9.33	51.15	6.98	.35	.05
English Proficiency: Yes	1,449	49.00	9.15	49.24	7.05	1,517	52.39	9.23	49.74	7.09	.37	.07
English Proficiency: No	28	44.02	7.72	47.43	6.74	30	47.00	8.14	46.27	6.64	.38	−.17

OPA
OFFICE OF PEOPLE ANALYTICS

Results

- Allowing calculators does not appear to alter the potential for adverse impact

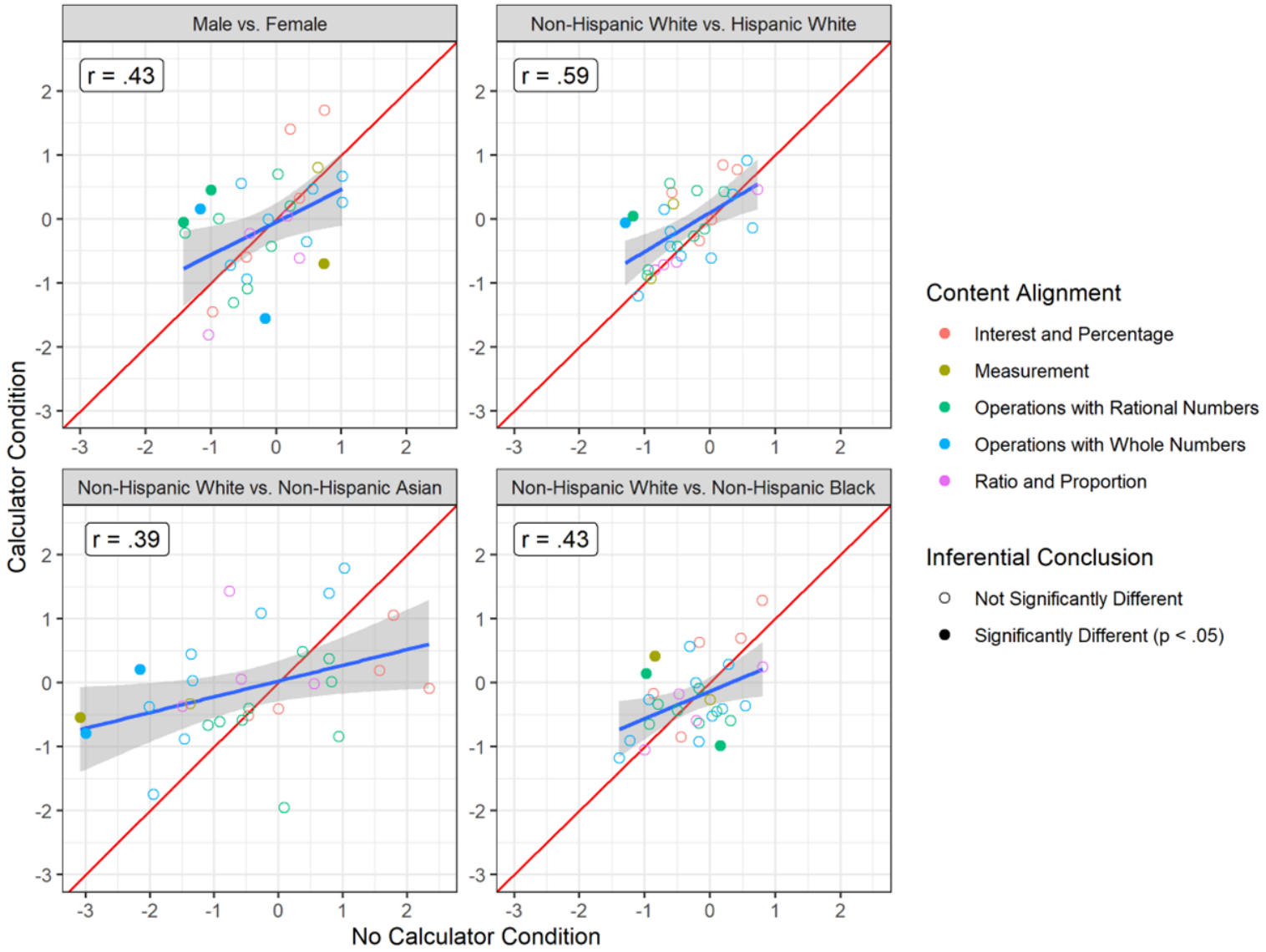
Subtest	Subgroup Contrast	Official Scores			Study Scores		
		d (SE) No Calculators	*d* (SE) Calculators	Difference	*d* (SE) No Calculators	*d* (SE) Calculators	Difference
AR	Male – Female	0.37 (.08)	0.40 (.08)	0.03 (.12)	0.37 (.08)	0.39 (.08)	0.02 (.12)
	Non-Hispanic White – Hispanic White	0.40 (.07)	0.41 (.07)	0.00 (.10)	0.44 (.07)	0.39 (.07)	−0.05 (.10)
	Non-Hispanic White – Non-Hispanic Asian	−0.21 (.15)	0.12 (.15)	0.33 (.19)	0.03 (.15)	0.19 (.15)	0.17 (.19)
	Non-Hispanic White – Non-Hispanic Black	0.83 (.07)	0.75 (.07)	−0.09 (.10)	0.77 (.07)	0.71 (.07)	−0.06 (.10)
	English Proficient – Not English Proficient	0.47 (.19)	0.30 (.19)	−0.17 (.27)	0.55 (.19)	0.58 (.19)	0.04 (.27)
MK	Male – Female	0.07 (.08)	0.01 (.08)	−0.07 (.12)	0.21 (.08)	0.14 (.08)	−0.07 (.12)
	Non-Hispanic White – Hispanic White	0.22 (.07)	0.22 (.07)	−0.00 (.10)	0.42 (.07)	0.37 (.07)	−0.05 (.10)
	Non-Hispanic White – Non-Hispanic Asian	−0.66 (.15)	−0.28 (.15)	0.38 (.20)	−0.20 (.15)	−0.06 (.15)	0.14 (.19)
	Non-Hispanic White – Non-Hispanic Black	0.35 (.07)	0.25 (.07)	−0.10 (.10)	0.52 (.07)	0.49 (.07)	−0.02 (.10)
	English Proficient – Not English Proficient	0.07 (.19)	0.32 (.19)	0.25 (.27)	0.26 (.19)	0.49 (.19)	0.23 (.27)

OPA
OFFICE OF PEOPLE ANALYTICS

Note: *d* indicates standardized mean difference between subgroups.

Results

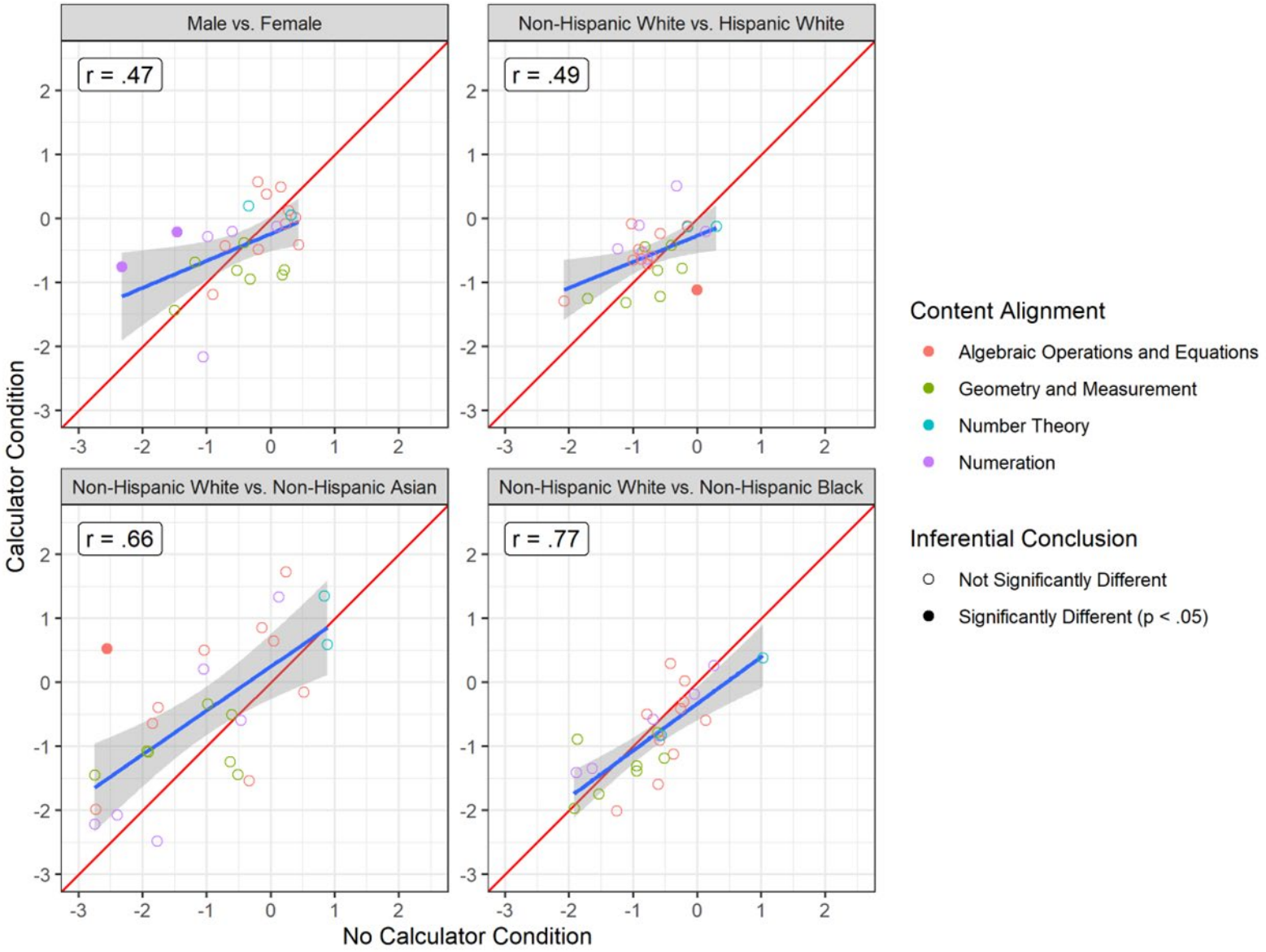
- Within condition DIF analysis (AR), significant differences in DIF between conditions were uncommon across subgroup comparisons



OPA
OFFICE OF PEOPLE ANALYTICS

Results (cont.)

- Within condition DIF analysis (MK), significant differences in DIF between conditions were uncommon across subgroup comparisons



OPA
OFFICE OF PEOPLE ANALYTICS

Summary & Conclusions for RQ3

- Calculators do not appear to differentially impact scores by demographic subgroups
  - Magnitude of between conditions standardized mean difference (*d)* is comparable across subgroups
  - Significant differences in DIF between conditions were uncommon across subgroup contrasts

OPA
OFFICE OF PEOPLE ANALYTICS

Research Question 4: Does calculator availability impact the amount of time needed to complete each math subtest?

- Mean differences in test times between conditions by subgroup

OPA
OFFICE OF PEOPLE ANALYTICS

Results (cont.)

Note: Bold numbers indicate statistically significant differences. Statistically non-significant moderate effect sizes are associated with small sample sizes.

Subgroup	No Calculator				Calculator				Effect Size	
	AR		MK		AR		MK		AR	MK
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*d*	*d*
Overall	31.06	9.86	13.19	5.63	28.26	9.25	13.25	5.66	**–0.29**	0.01
Female	32.05	9.60	13.68	5.97	29.92	9.44	14.60	6.36	**–0.22**	0.15
Male	31.00	9.88	13.16	5.66	28.08	9.16	13.09	5.54	**–0.31**	–0.01
Hispanic White	33.18	9.66	13.78	6.41	30.34	8.84	13.62	5.71	**–0.31**	–0.03
Non-Hispanic Asian	33.56	9.02	13.64	4.83	29.86	9.45	14.98	6.31	**–0.40**	0.23
Non-Hispanic Black	33.72	10.27	14.19	6.23	32.13	9.89	14.62	6.89	**–0.16**	0.07
Non-Hispanic White	28.66	9.18	12.60	4.94	24.97	7.77	12.22	4.62	**–0.43**	–0.08
English Proficiency: Yes	30.99	9.85	13.13	5.58	28.20	9.21	13.21	5.58	**–0.29**	0.01
English Proficiency: No	36.15	9.05	17.17	6.73	32.61	10.63	15.53	8.92	–0.36	–0.21

Summary & Conclusions for RQ4

- Calculators do not appear to differentially impact time spent by demographic subgroups
  - All subgroups completed AR more quickly when a calculator was available; the magnitude of the time spent difference was similar across subgroups
  - The impact of calculator availability on MK time spent was trivial to small for all subgroups

OPA
OFFICE OF PEOPLE ANALYTICS

Supplemental Analyses

- There were trivial to small differences between the No Calculator and Calculator conditions on some of the post-test questions. Participants in the Calculator condition reported feeling slightly more motivated and slightly less anxious than participants in the No Calculator condition.

Post-test Question	No Calculator			Calculator			
	M	*SD*	*n*	*M*	*SD*	*N*	*d*
How anxious were you while taking the two math tests today?	2.55	0.62	1,474	2.62	0.57	1,545	0.12
What was your motivation to answer questions correctly while taking these tests? (analysis sample)	1.44	0.50	1,485	1.39	0.49	1,557	-0.10
What was your motivation to answer questions correctly while taking these tests? (prior to data cleaning)	1.63	0.73	1,737	1.52	0.66	1,727	-0.16

OPA
OFFICE OF PEOPLE ANALYTICS

Note: For the anxiety question, the positive *d* value reflects higher anxiety in the No Calculator condition. For the motivation question, the negative *d* values reflect higher motivation in the Calculator condition.

General Conclusions and Implications

OPA
OFFICE OF PEOPLE ANALYTICS

Summary of Findings

- There is no discernible impact of allowing calculators on the factor structure or dimensionality of AR and MK
  - Parallel analysis, bifactor CFA analysis, and correlation analysis indicate no meaningful dimensionality differences between conditions for AR and MK
  - DTF results indicate some AR items are easier in Calculator condition
- Allowing calculators had no notable impact on item discrimination and subtest reliability

Summary of Findings (cont.)

- Some AR items were easier for participants in the Calculator condition than in the No Calculator condition; overall, AR scores were higher in the Calculator condition
  - Differences in AR test characteristic curves (TCCs) between conditions were minimal after using linear rescaling to account for the impact of calculators (overall impact of calculators on IRT parameters is primarily linear)
  - Scores of examinees who test with a calculator can be linked to the score scale of examinees who test without a calculator with a high degree of accuracy using linear transformations
    - **Note: this finding is likely limited to the specific, fixed linear forms used in this study, and we do not expect it to necessarily generalize to all P&P-ASVAB and CAT-ASVAB forms (see Limitations).**
- MK items tended not to be impacted by allowing calculators; overall, MK scores were not significantly different between conditions

Summary of Findings (cont.)

- The impact of allowing calculators is similar across demographic subgroups
  - Mean differences between the No Calculator and Calculator conditions were comparable across subgroups for both subtests
  - Where there were apparent differences across subgroups in potential performance gains in the Calculator condition, the subgroup sample sizes were small (meaning that sampling error cannot be ruled out as an explanation for the pattern of results observed)
- All subgroups completed AR more quickly when a calculator was available; this difference was statistically significant for all subgroups except non-English proficient participants
  - The numbers of non-English proficient participants were small for both the No Calculator and Calculator conditions, so this finding should be interpreted with caution
- There were no significant mean differences in testing times between conditions for MK

Limitations

- Study included only 30 AR and 25 MK items
  - A very small subset of the total inventory (approximately 10,000) of AR and MK items
  - It is possible the impact of calculators on other fixed-length, linear forms composed of different subsets of AR and MK items could be stronger or weaker than the current results
  - Other subtests that could be affected by calculator use, such as MC and EI, were not included
- Use of a fixed-length, linear form limits our ability to infer impact in CAT-ASVAB administrations, or even on other fixed-length, linear forms that may include a different mix of calculator-sensitive items
  - It seems reasonable to assume there will be a range across examinees in the number of calculator-sensitive items administered (i.e., some examinees might see significantly more calculator-sensitive items than other examinees)

Limitations (cont.)

- Use of a fixed-length, linear form limits our ability to infer impact in CAT-ASVAB administrations, or even on other fixed-length, linear forms that may include a different mix of calculator-sensitive items (cont.)
  - If calculators are permitted on the ASVAB, it will be important to account for the variability in calculator sensitivity across items to minimize the possibility that any given applicant could be advantaged or disadvantaged based on the number of calculator-sensitive items received
  - It would be inappropriate to apply a single scaling constant to all applicants provided with a calculator if some applicants receive fewer calculator-sensitive items than others
- All AR and MK item parameters, regardless of P&P or CAT format, would need to be rescaled based on a linkage of parameter estimates derived from larger samples of both examinees and items
- This rescaling would involve a universal scale transformation for item parameters on all forms, such that all item parameters for a given subtest would be adjusted via the same linear transformation, not form-specific transformations

Limitations (cont.)

- The P&P-ASVAB and CAT-ASVAB could be impacted by this universal rescaling in different ways
  - P&P-ASVAB forms, although psychometrically parallel at the time of their design, may contain different numbers of calculator (in)sensitive items
    - Variation in form-level calculator sensitivity could result in forms producing scores impacted by systematic biases, even after the average effect of calculators is taken into account
    - Forms with more calculator-sensitive items would produce overestimated scores, while forms with fewer calculator-sensitive items would produce underestimated scores
  - CAT-ASVAB forms could also be impacted by residual errors in parameter estimates after item parameters are rescaled, as those errors would impact the efficiency with which the CAT algorithm selects items

Implications if Calculator Use Is Allowed on AR and MK

- Psychometric implications
  - An equating study will be necessary to maintain statutorily-required AFQT qualification rates
    - USC, Title 10, Sec 520, mandates how AFQT is to be applied for the purpose of enlistment (statute mandates a limitation on enlistment of applicants with an AFQT score between 10 and 30)
    - This implies an ability to accurately estimate aptitude—allowing use of calculators on the ASVAB could result in changing the definition of the AFQT scores
  - Calculator use would affect both the CAT and P&P formats and multiple administration purposes (AFCT, PiCAT, VTest, ETP, etc.)
    - Will have implications for score scale as forms are recycled for different purposes
  - Between AR and MK, approximately 10,000 items have been developed, calibrated, and scaled under no-calculator conditions
    - All item parameters will need to be rescaled (a complementary study suggests relying on SME judgments of impact would be insufficient)

Implications if Calculator Use Is Allowed on AR and MK (cont.)

- Psychometric implications (cont.)
  - The linear transformation constants used to convert theta estimates to standard scores are based on linking form-specific score distributions to the 1997 Profile of American Youth (PAY97) norms under no-calculator conditions
    - These constants will need to be adjusted to account for calculator effects on score distributions
  - New specifications for item development would be needed to guide item writing for use on future ASVAB administrations if calculators are allowed
  - A new testing time would need to be determined to account for possible changes in the amount of time needed to complete AR, MK, and the remainder of the ASVAB

Implications if Calculator Use Is Allowed on AR and MK (cont.)

- Psychometric implications (cont.)
  - Even if equated, many uncertainties persist
    - Impact(s) on validity: decades of validity evidence is based on ASVAB administered without the use of calculators
    - There is also a potential concern of accurately assessing the ability of examinees at the high-end of AR achievement
      - Calculators could create a ceiling effect on AR for higher ability applicants such that the AR subtest may no longer be able to accurately measure/assess the ability of examinees at the high end of the ability distribution
    - We have or will have only some knowledge (a snapshot based on 30 AR & 25 MK items) of psychometric impacts on *difficulty*, *dimensionality*, *response time*, *fairness*, *norms*, and *composite cut scores*

Implications if Calculator Use Is Allowed on AR and MK (cont.)

- Logistic considerations
  - Determining when and how to distribute and collect calculators during ASVAB administrations
  - Distributing and maintaining calculators (including for overseas testing)
  - Distributing and transporting calculators for ASVAB CEP administrations
  - Determining who will provide and maintain calculators for each Service for Armed Forces Classification Test (AFCT) administrations
  - Addressing test security concerns associated with monitoring the use of the approved device (including the possibility that individuals might attempt to alter their calculator to use as a recording device)
  - Creating training/guidance for Test Administrators
    - Including guidance on enforcement of approved calculator
    - Determining if/how to prevent calculator use on non-math subtests (e.g., MC, EI)

Implications if Calculator Use Is Allowed on AR and MK (cont.)

- Practical considerations
  - Given the parallelism between conditions' equated TCCs, allowing calculators could put some examinees at a disadvantage if they choose not to make full use of the calculators
    - Choosing not to (consistently) use a calculator could reduce examinees' expected rates of correct responses (but they would be evaluated relative to calculator users)
    - Examinees who prefer not to use a calculator would effectively test under no-calculator conditions but be scored according to calculator-based standards
    - Scores would reflect a function of both math ability and individual differences in calculator use

Thank you!

For more information
please contact:

Kevin Bradley
kbradley@humrro.org
703.706.5647

OPA
OFFICE OF PEOPLE ANALYTICS

Appendix

OPA
OFFICE OF PEOPLE ANALYTICS

Demographic Characteristics of Analysis Sample*

Demographic	No Calculator		Calculator		Total		FY 2023 Applicants/Accessions	
	n	%	*n*	%	*n*	%	*n*	%
Female	158	10.6	169	10.9	327	10.8	80,986	25.1
Male	1,216	81.9	1,307	83.9	2,523	82.9	237,604	73.5
Data Not Available	111	7.5	81	5.2	192	6.3	4,653	1.4
Hispanic White	291	19.6	363	23.3	654	21.5	80,348	24.9
Non-Hispanic Asian	50	3.4	68	4.4	118	3.9	17,406	5.4
Non-Hispanic Black	326	22.0	313	20.1	639	21.0	87,395	27.0
Non-Hispanic White	629	42.4	652	41.9	1,281	42.1	113,921	35.2
Other†	63	4.2	63	4.0	126	4.1	16,317	5.1
Data Not Available	126	8.5	98	6.3	224	7.4	7,856	2.4

OPA
OFFICE OF PEOPLE ANALYTICS

*Shippers from 59 of 65 MEPS participated in the study. Demographic characteristics of sample was similar between conditions.

†Participants who provided ethnicity information and identified as American Indian, Alaska Native, Native Hawaiian, or Other Pacific Islander, and/or identified as Hispanic Black or Hispanic Asian.

Analysis Sample

Service	No Calculator		Calculator		Total		FY 23 Applicants/Accessions	
	n	%	*N*	%	*n*	%	*n*	%
Army	517	34.8	534	34.3	1,051	34.5	165,358	51.2
Air Force	306	20.6	285	18.3	591	19.4	56,736	17.6
Marine Corps	224	15.1	304	19.5	528	17.4	46,935	14.5
Navy	275	18.5	329	21.1	604	19.9	46,199	14.3
Coast Guard	47	3.2	32	2.1	79	2.6	6,679	2.1
Space Force*	13	0.9	0	0.0	13	0.4		
Invalid/Missing	103	6.9	73	4.7	176	5.8	1,336	0.4
Total	1,485	100.0	1,557	100.0	3,042	100.0	323,243	100.0

*Due to Space Force service code not yet being consistently implemented in data system, Space Force applicants are included with Air Force.