

SLIDES ONLY
NO SCRIPT PROVIDED

CLEARED
For Open Publication

May 08, 2024

Department of Defense
OFFICE OF PREPUBLICATION AND SECURITY REVIEW



Next Generation Testing: Building a Shared Vision and Roadmap

Kimberly Adams & Scott Oppler
Human Resources Research Organization

Briefing presented to the DACMPT
June 12, 2024

24-P-0551

Briefing Agenda

- Background Information
- Lessons Learned and Recommendations from Historical Review
- Evidence Consolidation and Synthesis
- Methodology Design and Pilot Test
- Next Steps
- Q&A

Background Information

What We Know

- Since 1976, the Armed Services Vocational Aptitude Battery (ASVAB) has been supporting the U.S. military's recruitment and accession objectives that ensure our Nation's security and defense readiness
- Defense Testing and Assessment Center (DTAC) and the Services have conducted numerous efforts over the past 5 decades to establish evidence for the ASVAB's:
 - Decisions about content changes (e.g., removing/adding subtests)
 - Selection and classification validity and efficiency
 - Psychometric prowess and fairness
 - Transition to computer adaptive testing (CAT) in 1996-1997
 - Implementation of unproctored CAT administration with verification testing (i.e., PiCAT/Vtest)
- ASVAB underwent a systematic review in 2005-2006, resulting in:
 - A number of changes to the Enlisted Testing Program (ETP) such as: implemented CAT at MET sites; increased seeding time for new items and measures; added tests of cyber knowledge and working memory to ASVAB platform; developed a non-verbal reasoning test (in progress)

Current Direction

- Given the complexities associated with making changes to the ASVAB, DTAC decided it was best to consider all new and existing tests at once, rather than on a case-by-case basis
- DTAC has been and continues to engage in efforts to evaluate the need for and viability of:
 - Adding new constructs and noncognitive measures
 - Modifying the subtests and/or the content comprising the battery
- DTAC seeks to understand ASAVB stakeholders' perspectives prior to specifying and/or implementing any changes
- DTAC seeks to leverage lessons learned from prior ASVAB change efforts to minimize obstacles and barriers to success

Objectives, Assumptions, and Considerations

Objectives	Assumptions	Considerations
<ul style="list-style-type: none">▪ Build a shared understanding of Next Generation Testing among ASVAB stakeholders▪ Leverage lessons learned from prior ASVAB change efforts to minimize obstacles and barriers▪ Develop a shared vision and roadmap for Next Generation Testing	<ul style="list-style-type: none">▪ ASVAB stakeholders have different perspectives about what assessments should be included in a future test battery and why▪ Consensus across ASVAB stakeholders is unrealistic▪ A shared understanding of perspectives among ASVAB stakeholders will inform the path forward	<ul style="list-style-type: none">▪ What evaluation factors are considered most important when selecting assessments for inclusion in a future test battery?▪ Which assessments are considered most important for inclusion in a future test battery?<ul style="list-style-type: none">• How do perspectives change based on an assessment's evaluation factor score?• How do perspectives change based on an assessment's impact on total testing time?

Overall Approach

- Identify lessons learned to minimize obstacles for gaining stakeholder acceptance of potential changes
- Consolidate and synthesize evidence to be used for evaluating assessments
- Design and pilot methodology for building a shared understanding on:
 - Relative importance of evaluation factors
 - Relative importance of assessments
 - Shared understanding of alternative test batteries
- Establish ASVAB Stakeholder Panel (ASP)
 - MAPWG members (or designated representative)
 - Other stakeholder representatives (e.g., IT, CEP)
- Implement methodology for identifying acceptable alternative test batteries
- Explore and evaluate alternative test batteries to inform ultimate decisions

Lessons Learned and Recommendations

Information Gathering Approaches

Historical Document Review

- Reviewed reports, meeting agendas, and notes, articles, and other documentation to understand ASVAB policy and content decisions and decision-making processes related to:
 - Enhanced Computer Administered Test (ECAT)
 - ASVAB change efforts from the late 1980s through the 1990s

Retrospective Interviews

- Conducted retrospective interviews with key individuals involved in prior change efforts to:
 - Fill gaps in information gathered during the historical document review
 - Gain better insight into processes used for building consensus among stakeholders and overcoming barriers
 - Gather retrospective thoughts on lessons learned about the processes

Historical Document Review: Key Contributors

- Defense Manpower Data Center (DMDC) Personnel Testing Division (PTD)
 - Currently DTAC under the Office of People Analytics (OPA)
 - Responsible for developing and maintaining the ASVAB
- Manpower Accession Policy Working Group (MAPWG)
 - Established in 1974, currently active, and composed of Service technical and policy representatives (voting members) and DTAC, AP, and MEPCOM representatives (non-voting members)
 - Forum for reviewing and discussing ongoing work pertaining to (1) research, development, implementation, and maintenance of accessions tests, and (2) military personnel selection and classification processes and policies
- Technical Advisory Selection Panel (TASP)
 - A short-term committee that was active 1989-1993 and composed of technical representatives from the Services and DMDC
 - Responsible for choosing tests for the ECAT battery for joint-service validation
- ASVAB Review Technical Committee (ART)
 - A subcommittee of the MAPWG that was active 1989-1993 and composed of the Services' technical representatives
 - Responsible for evaluating technical data on ASVAB subtests and other candidate assessments and making recommendations to the broader MAPWG

Historical Document Review: Research Summary

1980s	Late '80s to Early '90s	1990s
<ul style="list-style-type: none">▪ Validating the ASVAB for prediction of job performance in the Job Performance Measurement (JPM) projects▪ Developing and evaluating computerized cognitive, spatial, and psychomotor assessments to support occupation classification decisions▪ Supporting development of a Joint-Service ECAT	<ul style="list-style-type: none">▪ Evaluating potential impacts of transitioning from paper-and-pencil (P&P-ASVAB) to computer-adaptive (CAT-ASVAB) administration<ul style="list-style-type: none">• Expanding or modifying subtests• Exploring feasibility of adding ECAT tests by considering hardware, cost, operational use, and overall testing time	<ul style="list-style-type: none">▪ Evaluating ASVAB effectiveness for selection and classification▪ Evaluating each subtest's predictive and incremental validity, subgroup differences, and adverse impact on qualification rates▪ Evaluating impact of:<ul style="list-style-type: none">• Removing Coding Speed (CS) and Numerical Operations (NO)• Adding Assembling Objects (AO)

Historical Document Review: Key Findings

- Committee/panel recommendations were valued and accepted for ECAT decisions
- DMDC recommendations were evaluated by Services, with an emphasis on reaching full consensus for ASVAB decisions
 - Resulted in a 6-year impasse for decisions about CO, NO, and AO
- Deputy Assistant Secretary of Defense for Military Manpower Personnel Policy ultimately made decisions in 2000
 - CS became a special test available to the Services upon request
 - NO was dropped from the ASVAB
 - AO became an ASVAB subtest
 - AO was not administered to Career Exploration Program (CEP) students

Retrospective Interviews: Key Contributors

- Interviewed 7 key stakeholders involved during the 1990s, representing:
 - DoD Accession Policy
 - DMDC/PTD
 - MAPWG and other representatives from the Air Force, Army, and Navy
- Asked a standard list of questions about ECAT and ASVAB decisions
 - Who was involved in the decision-making process?
 - How were decisions made?
 - What barriers existed to making changes and what advice is there to overcome them (ASVAB only)?
 - How could the decision-making process be improved, retrospectively?

Retrospective Interviews: Key Findings

ECAT Decision Processes

- Included representatives beyond MAPWG
- Used a collaborative, psychometric-oriented evaluation approach to reach consensus on decisions
- Focused on validity and reliability and addressing psychometric research needs
- No improvements to ECAT decision-making processes suggested

ASVAB Decision Processes

- Included representatives beyond MAPWG
- Involved independent research teams with:
 - Different types/sources of criterion data resulted in discrepant findings
 - Unequal reliance on adverse impact results
- Impacted by Services' constraints:
 - Lacked validity evidence to inform changes
 - Required IT system modifications
- Several improvements to ASVAB decision-making processes suggested

Lessons Learned: Approaches that Hinder Decision Making

- Failing to take the time to:
 - Understand how stakeholders are weighing the importance of psychometric parameters
 - Establish clear, consistent metrics and guidance for comparatively evaluating findings
 - Understand the logistical factors affecting different options early enough in the process
- Initiating the process without an expectation for making compromises
 - Expecting unanimous agreement across stakeholders
 - Engaging in research focused solely on a single stakeholder's perspective
- Expecting advisory groups to identify solutions for resolving differences
- Waiting too long to engage with high-ranking personnel with authority to support decisions and allocate resources

Lessons Learned: Approaches that Help Decision Making

- Encourage collaborative research and decision-making processes
 - Establish group(s) to support collaboration
 - Include technical and policy stakeholder representation on group(s)
 - Create continuity of key members across groups
- Identify potential challenges and constraints for changes early in the process
 - Engage with IT and other support services early to allow enough time for changes
- Anchor rationales for changes to specified evaluation criteria (i.e., scorecard)
- Consider alternative configurations of ASVAB and special tests
 - Address ownership/maintenance of special tests
- Establish an advisory committee with high-ranking personnel maintaining authority for policy decisions and resource allocation

Lessons Learned: Strategic Approach

Actions for Building a Shared Vision

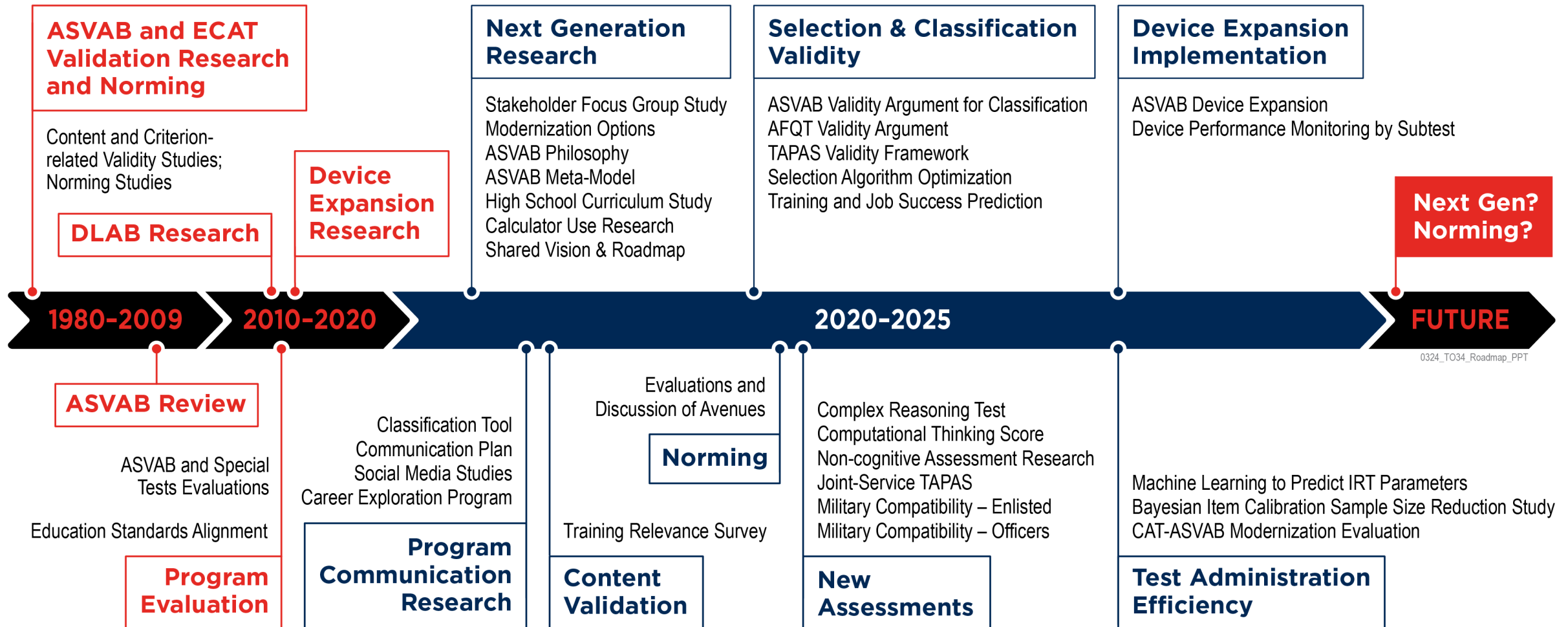
1. Establish a stakeholder engagement plan
2. Seek a shared understanding for an integrated vision
3. Set expectation for compromises rather than consensus
4. Understand and evaluate all potential challenges and constraints early in the process
5. Engage with IT and other support services early enough to implement changes
6. Engage with high-ranking personnel with authority for policy and resource decisions
7. Seek research recommendations, not resolutions, from DACMPT and other advisors

Evidence Consolidation and Synthesis

Consolidation and Synthesis Approach

- Gathered and reviewed relevant research, including but not limited to:
 - ASVAB Evaluation results presented at Sept 2020 DAC Meeting
 - Psychometric checklists for ASVAB subtests and special tests
 - ASVAB stakeholder focus group efforts
 - AFQT, ASVAB, and TAPAS validity arguments
 - Content and criterion-related validity evidence
 - ASVAB meta-model for predicting success in military training and on the job
 - ASVAB philosophy
- Identified factors to be evaluated as part of the methodology
- Selected assessments for evaluation in the methodology
 - Excluded TAPAS and Computational Thinking given requirement for pre-enlistment testing
- Documented evidence for each assessment on each evaluation factor (i.e., Evaluation Matrix)

Research and Evaluation Efforts by Decade for Next Generation Testing



0324_TO34_Roadmap_PPT

Evaluation Factors

Statistical & Psychometric Elements

- Reliability
- Selection Validity
- Classification Validity
- Subgroup Differences
- Resistance to:
 - Compromise (cheating/faking)
 - Practice effects
 - Hardware effects
 - Mode (CAT vs P&P)
 - Device familiarity
 - Obsolescence
 - Local dependence

Administrative & Policy Considerations

- Test Administration Time Feasibility
- Test Development and Maintenance Costs
- Opportunity for Examinees to Learn Content in High School Curriculum or Extracurricular Activities
- Usefulness for Selecting Candidates for Foreign Language Training
- Usefulness to the Career Exploration Program

Assessments for Consideration in Future Battery

AFQT Subtests

- Arithmetic Reasoning (AR)
- Mathematics Knowledge (MK)
- Paragraph Comprehension (PC)
- Word Knowledge (WK)

ASVAB Subtests

- Auto Information (AI)
- Shop Information (SI)
- Auto/Shop Information (AS)
- Electronics Information (EI)
- General Science (GS)
- Mechanical Comprehension (MC)
- Assembling Objects (AO)

Special & Experimental Tests

- Cyber Test (CT)
- Coding Speed (CS)
- Mental Counters (MCt)
- Complex Reasoning (CR)
- DLAB-Biographical Information (DLAB-Bio)

Prior Evaluation Effort of ASVAB Contents

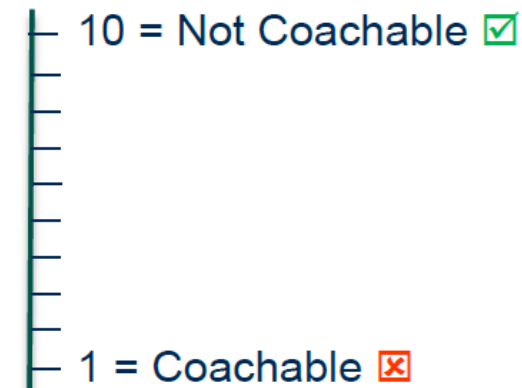
- Leveraged DTAC's evaluation ratings of the ASVAB subtests conducted in 2019-2020
 - Step 1: Reviewing the history of current ASVAB tests and why they were originally included in the battery
 - Step 2: Completing psychometric checklists and evaluating the psychometric value/limitations for each test
 - Step 3: Evaluating the usefulness/appropriateness of existing tests with the current applicant population
 - Step 4: Evaluating the item/form development costs
 - Step 5: Evaluating the ease/difficulty of developing good quality items
 - Step 6: Evaluating the durability of test content
 - Step 7: Evaluating the appropriateness/efficiency of content coverage across tests
 - Step 8: Evaluating the vulnerability of content to compromise and other unwanted effects
 - Step 9: Evaluating the efficiency of each test
 - Step 10: Evaluating the psychometric impact of shortening or combining various tests
 - Step 11: Evaluating the psychometric impact of dropping various tests

Example Rating from ASVAB Evaluation

- Based on prior research, DTAC assigned an assessment score of 1 to 10, with 10 being the best score, on each evaluation factor

Task 9a: Coachability

Subtest	Coachability Rating
General Science	10
Arithmetic Reasoning	10
Word Knowledge	9
Paragraph Comprehension	8
Mathematics Knowledge	10
Electronics Information	10
Automotive Information	10
Shop Information	10
Mechanical Comprehension	8
Assembling Objects	5



See slide 60 from 10a-DPAC_DAC_NextGenTesting_09-18-2020

Additional Evaluation Factor Ratings

- Added special and experimental tests to evaluation plan
 - DTAC applied similar approach to generate an evaluation factor rating for each special and experimental test and for AI and SI separately
- Added selection validity and classification validity as evaluation factors
 - HumRRO applied a similar approach with internal validity experts that have extensive knowledge of military testing to generate factor ratings for all assessments
- Evaluation Matrix includes evaluation factor rating for each assessment
 - Used Evaluation Matrix to generate a “scorecard” for each assessment

Evaluation Matrix Snapshot

Evaluation Factor	AR	MK	PC	WK	AI	SI	EI	GS	MC	AO	CT	CS	Mct	CR	DLAB -Bio
Reliability	8	8	6	10	10	8	8	8	8	8	4	4	10	8	NA
Resistance to:															
• Compromise	6	6	7	2	7	7	8	7	8	10	8	10	10	9	1
• Coaching	10	10	8	9	10	10	10	10	8	5	10	10	10	9	2
• Practice	8	6	10	10	7	7	6	6	10	10	10	6	6	10	10
• Hardware	6	6	10	10	10	10	10	10	5	4	10	2	2	3	10
• Mode	10	8	5	10	6	6	5	10	10	9	9	5	1	7	10
• Device Familiarity	6	7	8	10	8	8	8	6	8	6	8	5	5	6	10
• Obsolescence	10	10	10	10	3	8	6	7	10	10	1	10	10	10	10
• Local Dependence	9	5	9	9	7	7	7	7	5	10	7	10	10	10	NA
Selection Validity	10	7	8	9	4	4	6	6	6	6	3	5	4	7	NA
Classification Validity	7	5	5	4	7	7	7	7	8	5	7	3	5	6	7

Example Scorecard for Arithmetic Reasoning Subtest

Psychometrics

RELIABILITY	SCORE	MEDIAN OF ASVAB SUBTESTS
Extent to which the test yields scores that are consistent/reproducible	8	8
<i>Notes:</i> Scores are based on a 10-point Likert scale where 10 = Very strong reliability; 8 = Strong reliability; 6 = Moderate reliability; 4 = Weak reliability; and 1 = Unacceptable level of reliability.		

RESISTANCE TO SOURCES OF ERROR	SCORE	MEDIAN OF ASVAB SUBTESTS
Resistance to vulnerabilities from:		
a. Compromise	6	7
b. Coaching effects	10	10
c. Practice effects	8	8
d. Hardware effects	6	10
a. Mode (P&P vs. CAT) effects	10	9
b. Effects due to device familiarity	6	8
c. Items becoming dated (obsolesce)	10	10
d. Local dependence in test content	9	7
<i>Notes:</i> Scores are based on a 10-point Likert scale where 10 = high resistance and 1 = low resistance.		

VALIDITY	SCORE	MEDIAN OF ASVAB SUBTESTS
Selection Validity: Content or criterion-related evidence of validity across all occupations	10	6
Classification Validity: Content or criterion-related evidence that the assessment differentiates occupations (e.g., extent of current classification use)	7	7
<i>Notes:</i> Scores are based on a 10-point Likert scale where 10 = Very strong validity evidence; 8 = Strong validity evidence; 6 = Moderate validity evidence; 4 = Weak validity evidence; and 1 = Little or no validity evidence.		

Subgroup Differences

SUBGROUP	SCORE	MEDIAN OF ASVAB SUBTESTS
Minimal differences among racial/ethnic/gender subgroup scores	5	4
<i>Notes:</i> Scores are based on a 10-point Likert scale where 10 = Very small differences; 8 = Mostly small differences; 6 = Small to moderate differences; 4 = Moderate differences; and 1 = Large differences.		

Operational Considerations

OPERATIONAL CONSIDERATION	SCORE	MEDIAN OF ASVAB SUBTESTS
Testing time	55 minutes ^a	18 minutes
Development and maintenance costs	9 ^b	6
<i>Notes:</i> ^a Testing time reflects the allocated time (in minutes) without seed (tryout) items. ^b Scores are based on a 10-point Likert scale where 10 = Very cost-efficient; 8 = Cost-efficient; 6 = Moderately costly; 4 = Costly; and 1 = Very costly.		

Other Considerations

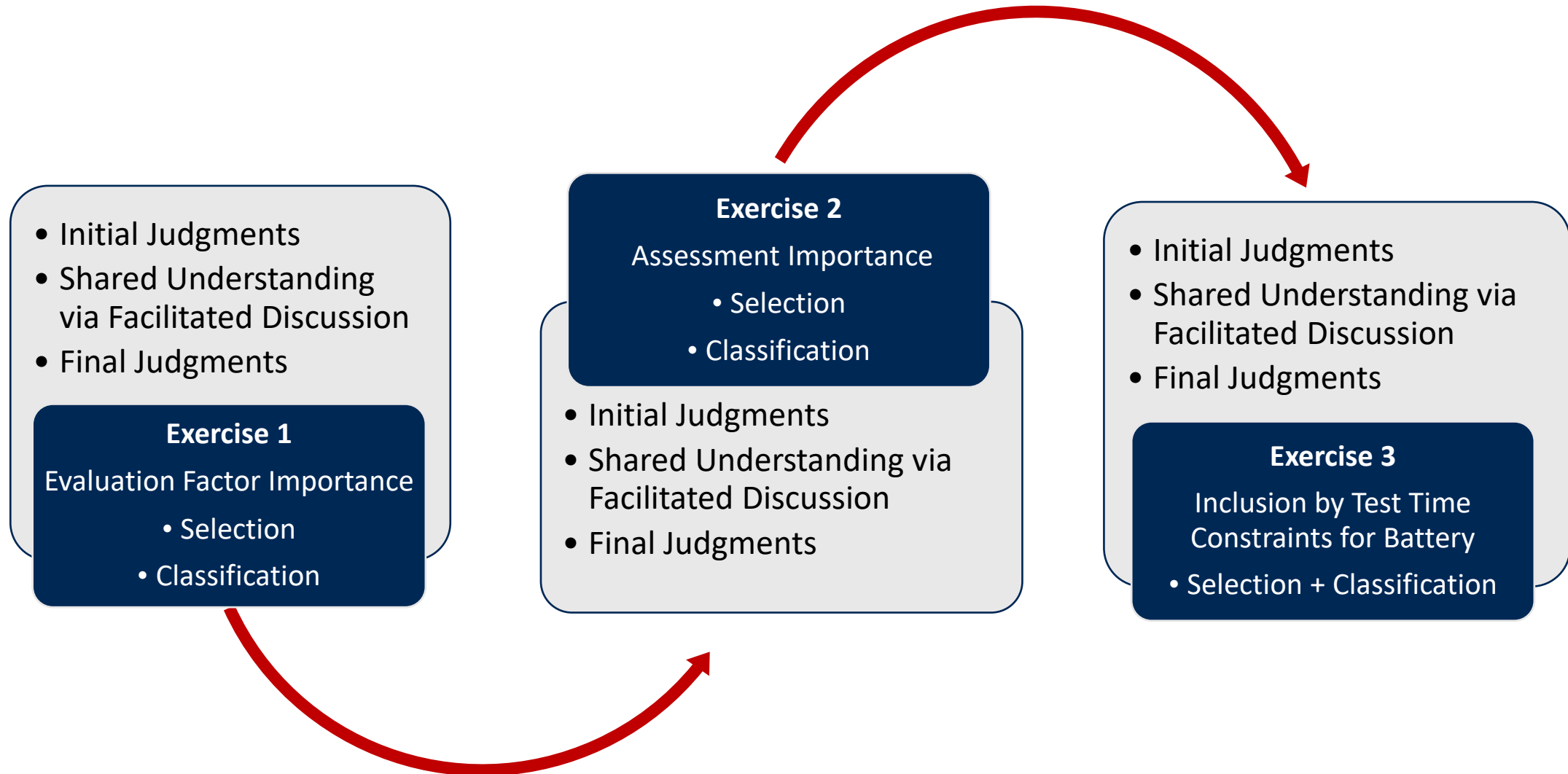
OTHER CONSIDERATIONS	SCORE	MEDIAN OF ASVAB SUBTESTS
Opportunity for applicant population to learn test content through high school curriculum	TBD ^a	
Extent to which the assessment is useful in selecting candidates for foreign language training	5 ^b	2
Compatibility with the CEP	TBD	
Compatibility with device expansion plans	TBD	
<i>Notes:</i> ^a The HS curriculum study is near completion with results likely to be available in March 2024. ^b Scores are based on a 10-point Likert scale where 10 = Very strong; 8 = Strong; 6 = Moderate; 4 = Weak; and 1 = Little or no usefulness.		

Methodology Design and Pilot Test

Overview of Methodology

- Data collection methodology with ASVAB Stakeholder Panel (ASP)
 - Exercise 1: Collect judgments of relative importance of the evaluation factors (separately for selection and classification)
 - Exercise 2: Collect judgments of relative importance of the assessments based on Exercise 1 results applied to scorecards (separately for selection and classification)
 - Exercise 3: Collect judgments of relative importance of including each assessment in a future test battery (for selection and classification combined) under different administration time constraints (e.g., 150 mins, 180 mins, 210 mins)

Overview of Exercises



Proposed ASVAB Stakeholder Panel (ASP)

MAPWG DoD Representatives	MAPWG Technical and Policy Representatives	Other Stakeholder Group Representatives
<ul style="list-style-type: none">▪ OPA/DTAC (2)▪ Accession Policy (1)▪ MEPCOM (2)	<ul style="list-style-type: none">▪ Air Force (2)▪ Army(2)▪ Coast Guard (1)▪ Marine Corps (2)▪ Navy (2)▪ Space Force (2)	<ul style="list-style-type: none">▪ Recruiters (4 total from AF/SF, Army, MC, Navy)▪ Classifiers (2 total from Army, Navy)▪ Trainers (2 total from AF/SF, MC)▪ MEPCOM Education Service Specialists (1)▪ IT System Representatives<ul style="list-style-type: none">• Services (5 total from AF/SF, Army, CG, MC, Navy)• MEPCOM (1)

Pilot Test Objectives and Procedures

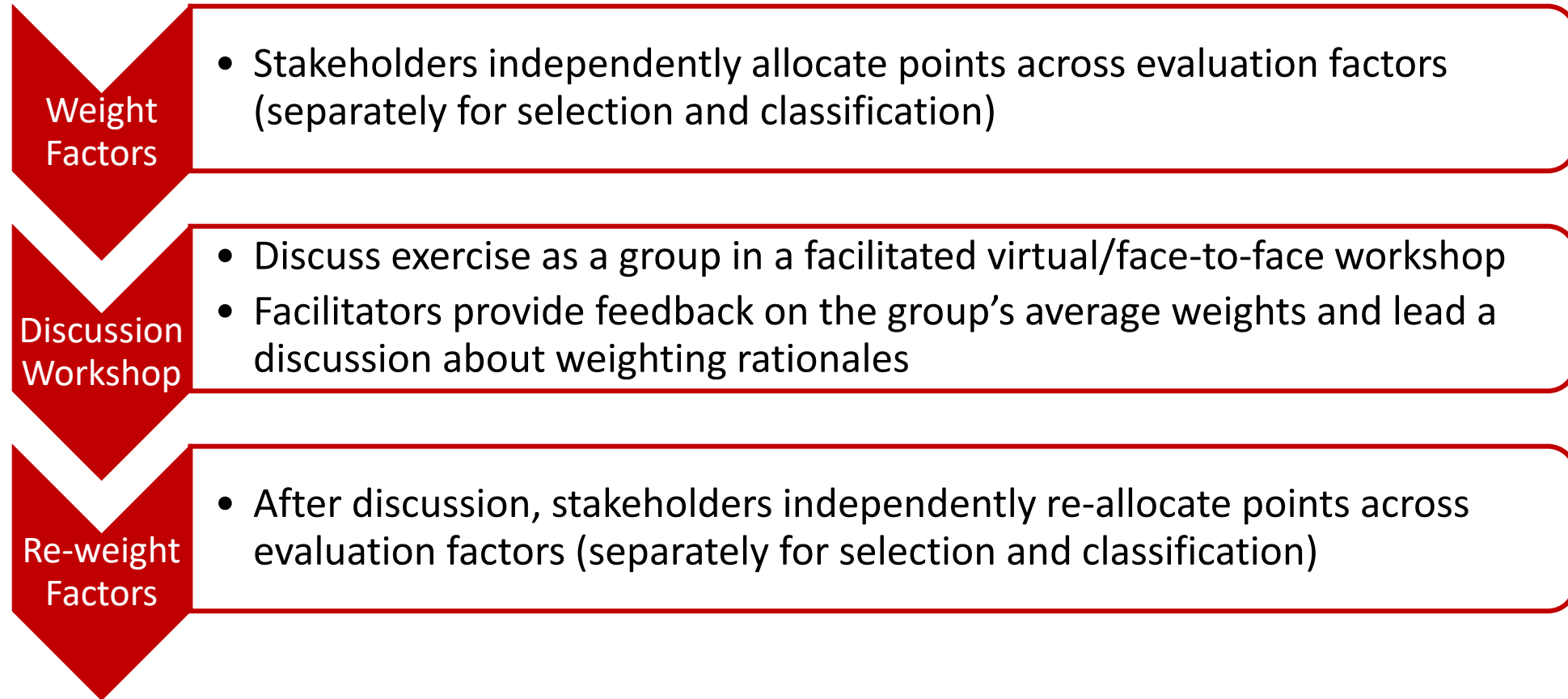
■ Objectives

- Try out different judgment methods (importance ratings vs. point allocations)
- Conduct usability testing on data collection forms
- Confirm clarity of instructions and feedback provided on judgments
- Practice facilitating discussion on a subset of evaluation factors
- Verify results provide type of information needed

■ Procedures

- Implemented each exercise virtually with pilot test participants
- HumRRO participants assigned to role play a stakeholder group based on their prior knowledge of that group (used stakeholder focus group results to supplement understanding)
- DTAC participants completed exercises as themselves

Exercise 1: Relative Importance of Evaluation Factors



Exercise 1 Feedback Report on Initial Judgments

Evaluation Factor	Selection				Evaluation Factor	Classification			
	Your Allocation	Round 1 Group Mean	Round 1 Standard Deviation	Difference from Round 1 Group Mean		Your Allocation	Round 1 Group Mean	Round 1 Standard Deviation	Difference from Round 1 Group Mean
Reliability and Psychometric Elements (RPE)					Reliability and Psychometric Elements (RPE)				
Reliability - the extent to which the test yields scores that are consistent/reproducible.	7	13	7	-6	Reliability - the extent to which the test yields scores that are consistent/reproducible.	7	13	7	-6
Resistance of item content and item pools to compromise (e.g., cheating/faking)	5	4	2	1	Resistance of item content and item pools to compromise (e.g., cheating/faking)	4	4	2	0
Resistance to coaching (i.e., score gains when coached)	5	3	2	2	Resistance to coaching (i.e., score gains when coached)	4	3	2	1
Resistance to practice effects (i.e., score gains with practice)	5	3	2	2	Resistance to practice effects (i.e., score gains with practice)	4	3	2	1
Resistance to hardware effects (i.e., score change across hardware such as monitor size)	5	3	2	2	Resistance to hardware effects (i.e., score change across hardware such as monitor size)	4	3	2	1
Resistance to mode (CAT vs. paper and pencil) effects	5	3	2	2	Resistance to mode (CAT vs. paper and pencil) effects	4	3	2	1
Resistance to device familiarity effects	5	2	1	3	Resistance to device familiarity effects	4	2	1	2
Resistance to obsolescence (i.e., items becoming dated)	5	4	4	1	Resistance to obsolescence (i.e., items becoming dated)	6	4	3	2
Resistance to local dependence (e.g., items that have content that informs other items)	5	2	3	3	Resistance to local dependence (e.g., items that have content that informs other items)	5	2	3	3
RPE Category Total	47	37	13	10	RPE Category Total	42	37	13	5
Selection Validity					Selection Validity				
Content or criterion-related evidence of validity across all occupations	7	17	13	-10	Content or criterion-related evidence of validity across all occupations	3	7	6	-4
Classification Validity					Classification Validity				
Content or criterion-related evidence that the assessment differentiates occupations (e.g., extent of current classification use)	3	5	6	-2	Content or criterion-related evidence that the assessment differentiates occupations (e.g., extent of current classification use)	8	20	13	-12
Subgroup Differences					Subgroup Differences				
Minimal differences among racial/ethnic/gender subgroup scores	6	8	3	-2	Minimal differences among racial/ethnic/gender subgroup scores	6	7	4	-1
Administration Considerations (AC)					Administration Considerations (AC)				
Administration Time Feasibility - Assessments that are high on administration time feasibility can be administered in a short amount of time.	7	6	5	1	Administration Time Feasibility - Assessments that are high on administration time feasibility can be administered in a short amount of time.	7	5	4	2
Development and maintenance cost - Low cost of item development (including ease/cost of developing good items, potential for using automated item generation, and the number of new items needed to replenish item pools)	5	4	2	2	Development and maintenance cost - Low cost of item development (including ease/cost of developing good items, potential for using automated item generation, and the number of new items needed to replenish item pools)	5	4	2	1
AC Category Total	12	10	5	2	AC Category Total	12	9	4	3
Policy Considerations (PC)					Policy Considerations (PC)				
Opportunity for applicant population to learn test-related knowledge, skill, or ability in high school or extracurricular	7	7	8	0	Opportunity for applicant population to learn test-related knowledge, skill, or ability in high school or extracurricular activities	8	5	3	3
Extent to which the assessment is useful in selecting candidates for foreign language training	3	3	1	0	Extent to which the assessment is useful in selecting candidates for foreign language training	3	3	2	0
Extent to which the assessment is useful for the CEP program	3	6	4	-3	Extent to which the assessment is useful for the CEP program	3	4	4	-1
PC Category Total	13	15	11	-2	PC Category Total	14	12	6	2

Exercise 1 Applied Feedback Report from Final Judgments

Importance of Assessments Based on Scorecard Data and Your Exercise 1 Allocations

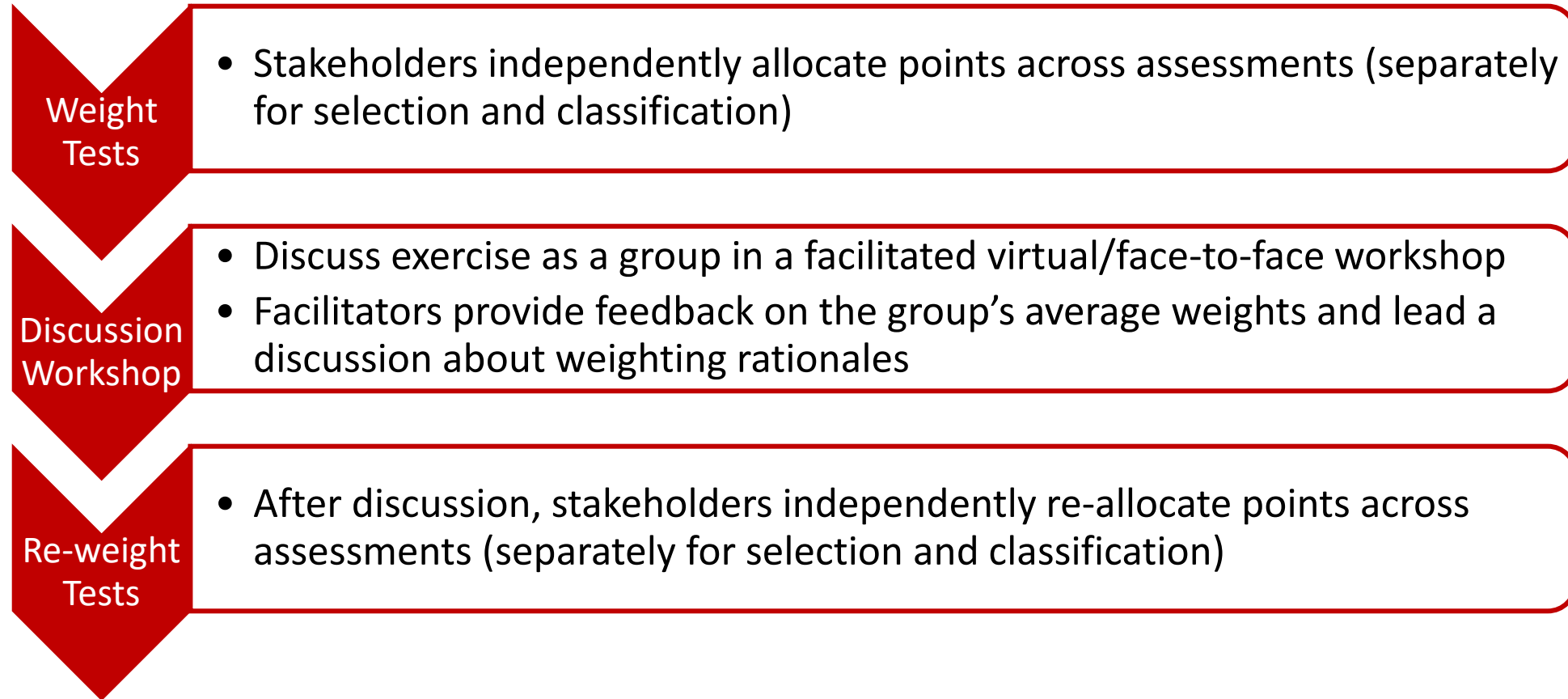
	AR	MK	PC	WK	AI	SI	EI	GS	MC	AO	CT	CS	Mct	CR	DLAB Bio
Empirical Point Allocation	8.3405	7.5541	7.4544	7.9307	5.6545	5.5881	6.6680	6.7012	6.6902	7.5320	5.5714	6.7566	5.9591	7.8532	3.7458
Rank	1	4	6	2	12	13	10	8	9	5	14	7	11	3	15

Empirical Point Allocation



Evaluation Factor	AR	MK	PC	WK	AI	SI	EI	GS	MC	AO	CT	CS	Mct	CR	DLAB Bio
Reliability and Psychometric Elements (RPE)															
Reliability - the extent to which the test yields scores that are consistent/reproducible	8	8	6	10	10	8	8	8	8	8	4	4	10	8	NA
Resistance of item content and item pools to compromise (e.g., cheating/faking)	6	6	7	2	7	7	8	7	8	10	8	10	10	9	1
Resistance to coaching (i.e., score gains when coached)	10	10	8	9	10	10	10	10	8	5	10	10	10	7	2
Resistance to practice effects (i.e., score gains with practice)	8	6	10	10	7	7	8	10	8	8	10	6	8	8	10
Resistance to hardware effects (i.e., score change across hardware such as monitor size)	6	6	10	10	10	10	10	10	5	4	10	2	2	3	10
Resistance to mode (CAT vs. paper and pencil) effects	10	8	5	10	6	6	5	10	10	9	9	5	1	7	10
Resistance to device familiarity effects	6	7	8	10	8	8	8	6	8	6	8	5	5	6	10
Resistance to obsolescence (i.e., items becoming dated)	10	10	10	10	3	8	6	7	10	10	1	10	10	10	10
Resistance to local dependence (e.g., items that have content that informs other items)	9	5	9	9	7	7	7	7	5	10	7	10	10	10	NA
Selection Validity															
Content or criterion-related evidence of validity across all occupations	10	7	8	9	4	4	6	6	6	6	3	5	4	7	NA
Classification Validity															
Content or criterion-related evidence that the assessment differentiates occupations (e.g., extent of current classification use)	7	5	5	4	7	7	7	7	8	5	7	3	5	6	7
Subgroup Differences															
Minimal differences among racial/ethnic/gender subgroup scores	5	8	6	4	3	2	4	4	4	9	7	9	4	9	NA
Administration Considerations (AC)															
Administration Time Feasibility - Assessments that are high on administration time feasibility can be administered in a short amount of time.	1	5	6	10	10	10	10	9	7	8	9	10	5	5	10
Development and maintenance cost - Low cost of item development (including ease/cost of developing good items, potential for using automated item generation, and the number of new items needed to replenish item pools)	9	8	2	7	3.5	3.5	6	5	4	6	6	6	6	6	10
Policy Considerations (PC)															
Opportunity for applicant population to learn test-related knowledge, skill, or ability in high school or extracurricular activities	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Extent to which the assessment is useful in selecting candidates for foreign language training	5	8	9	10	1	1	1	2	2	2	3	9	9	9	10
Extent to which the assessment is useful for the CEP program	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Exercise 2: Relative Importance of Assessments



Exercise 2 Feedback Report (Initial or Final Judgments)

Assessment		Selection			Assessment		Classification		
	Your Allocation	Initial Round Group Mean	Initial Round Standard Deviation	Difference from Initial Round Group Mean		Your Allocation	Initial Round Group Mean	Initial Round Standard Deviation	Difference from Initial Round Group Mean
Current AFQT Subtests					Current AFQT Subtests				
Word Knowledge (WK)	27	13	5	14	Word Knowledge (WK)	7	8	3	-1
Paragraph Comprehension (PC)	23	12	4	11	Paragraph Comprehension (PC)	8	8	2	0
Mathematics Knowledge (MK)	26	12	5	14	Mathematics Knowledge (MK)	7	8	2	-1
Arithmetic Reasoning (AR)	24	13	4	11	Arithmetic Reasoning (AR)	9	9	2	0
<i>Current AFQT Subtests Total</i>	100	50	17	50	<i>Current AFQT Subtests Total</i>	31	33	7	-2
Other ASVAB Subtests					Other ASVAB Subtests				
General Science (GS)	0	7	3	-7	General Science (GS)	9	7	3	2
Electronics Information (EI)	0	5	3	-5	Electronics Information (EI)	10	9	3	1
Mechanical Comprehension (MC)	0	5	3	-5	Mechanical Comprehension (MC)	10	9	2	1
Auto and Shop Information (AS)	0	4	2	-4	Auto and Shop Information (AS)	10	10	2	0
Assembling Objects (AO)	0	7	4	-7	Assembling Objects (AO)	8	6	2	2
<i>Other ASVAB Subtests Total</i>	0	29	10	-29	<i>Other ASVAB Subtests Total</i>	47	40	4	7
Special and Experimental Tests					Special and Experimental Tests				
Cyber Test (CT)	0	4	3	-4	Cyber Test (CT)	6	9	2	-3
Coding Speed (CS)	0	5	2	-5	Coding Speed (CS)	6	6	3	0
Complex Reasoning (CR)	0	6	3	-6	Complex Reasoning (CR)	5	6	3	-1
Mental Counters (MCT)	0	6	4	-6	Mental Counters (MCT)	5	5	2	0
<i>Special and Experimental Tests Total</i>	0	21	8	-21	<i>Special and Experimental Tests Total</i>	22	27	6	-5

Exercise 3: Inclusion of Assessment in Future Battery

Inclusion Selections

- Stakeholders independently select assessments for inclusion in a future battery to be used for selection and classification, under three different test time scenarios

Discussion Workshop

- Discuss exercise as a group in a facilitated virtual/face-to-face workshop
- Facilitators provide feedback on the group's most commonly chosen assessments and lead a discussion about assessment selection rationales

Re-select Inclusions

- After discussion, stakeholders independently select assessments for inclusion in a future battery to be used for selection and classification, under the same three test time scenarios

Exercise 3 Feedback Report (Initial and Final Judgments)

Assessments	2.5 hours		3 hours		3.5 hours		CAT-ASVAB Time Allotted (minutes)
	Your Selection	% of Group Including Assessment	Your Selection	% of Group Including Assessment	Your Selection	% of Group Including Assessment	
Current AFQT Subtests							
Arithmetic Reasoning (AR)	x	60%	x	59%	x	70%	55
Mathematics Knowledge (MK)	x	85%	x	56%	x	100%	31
Paragraph Comprehension (PC)	x	96%	x	79%	x	80%	27
Word Knowledge (WK)		2%		6%		0%	9
Other ASVAB Subtests							
Auto Information (AI)	x	100%	x	63%	x	40%	6
Shop Information (SI)		13%	x	80%	x	30%	7
Electronics Information (EI)	x	100%	x	35%	x	100%	10
General Science (GS)		16%		5%		60%	12
Mechanical Comprehension (MC)	x	100%	x	36%	x	100%	22
Assembling Objects (AO)		40%		3%		30%	18
Special and Experimental Tests							
Cyber Test (CT)		20%	x	40%	x	100%	12
Coding Speed (CS)		24%		3%		10%	7
Mental Counters (MCT)		35%		24%		20%	11
Complex Reasoning (CR)		10%		58%	x	80%	35
DLAB Biographical Inventory		2%	x	51%	x	100%	5
Total Time Allocated							
Maximum Time Allocation	150 minutes	150 minutes	180 minutes	180 minutes	210 minutes	210 minutes	
<i>Your Allocation</i>	151 minutes	151 minutes	175 minutes	166 minutes	210 minutes	209 minutes	

Next Steps

MAPWG Input Needed

- Input on ASP representatives
- Input on Implementation Plan
 - Determine feasibility of in-person workshop with half-day virtual introduction
 - Availability of FY24 funds to support travel
 - Feasibility of adding travel to FY25 budget, if necessary
 - Ability to share location from which ASP participant would be traveling
 - Determine willingness to support a fully virtual workshop, if necessary
 - Feasibility of assigning independent judgments as homework
 - Feasibility of conducting each exercise in one day (e.g., two 4-hour sessions separated by a lunch break) or over two or more consecutive days
 - Determine timeframe for implementation with ASP

Further Consideration of Alternative Test Batteries

- Process will not end with identification of potentially acceptable alternative test batteries from Exercise 3
- Need to design an approach for further consideration of alternative test batteries
 - May involve establishing an advisory committee with high-ranking Service personnel with authority for policy decisions and resource allocation
 - May require future research to inform decisions

Q&A

Guidance from the DAC

- Feedback on the methodology?
- Suggestions on representatives to the ASP?
 - Guidance on finalizing representative participants?
- Guidance on advisory committee approach for leveraging the shared vision to make final decisions on a future test battery for Next Generation Testing?
 - Necessary?
 - Factors to consider?
 - Other approaches to consider?
- Suggestions for future steps/research?

Thank you!

For more information
please contact:

Kimberly Adams
kadams@humrro.org
703.236.4303