DEFENSE ADVISORY COMMITTEE ON MILITARY PERSONNEL TESTING

December 15-16, 2022 Meeting

Office of the Under Secretary of Defense (Personnel and Readiness)

Minutes approved for public release.

Nancy L. Tippins March 14, 2023

Dr. Nancy Tippins, Chair, DACMPT **DATE**

DEFENSE ADVISORY COMMITTEE
ON
MILITARY PERSONNEL TESTING

December 15-16, 2022

The Fiscal Year (FY) 2023 first session of the Defense Advisory Committee on Military Personnel Testing (DACMPT) was held at the Hyatt Regency Monterey, Monterey, CA on December 15-16, 2022. The meeting was conducted in person; however, one DACMPT committee member and several briefers participated virtually using the Microsoft® Teams online collaboration tool. Dr. Sofiya Velgach (Assistant Director, Office of Accession Policy [AP]) opened the meeting by stating that it was being held under the provisions of the Federal Advisory Committee Act (FACA) of 1972 (5 USC, Appendix, as amended), the government in the Sunshine Act of 1976 (5 USC, 552b, as amended), and all other governing Federal statutes and regulations, and open to the public. She said the meeting agenda was available on the DACMPT website[1] and public comments would be received at the end of each day's scheduled sessions.

Dr. Velgach thanked the committee members for their participation and the presenters for their support of the committee's activities. She then introduced the inbound Director of AP, Dr. Katherine Helland. Dr. Helland spoke briefly about her experience in recruiting market research and commented on the legacy of the Armed Services Vocational Aptitude Battery (ASVAB). Dr. Helland remarked that this year marks the 50th anniversary of the all-volunteer force before adding that it also presents the most challenging recruiting environment to date. She explained that over half of the youth have not considered joining the Armed Services and added that some key constituencies were questioning if a standardized test for military entrance is still needed. Dr. Helland told the committee members that she was interested in hearing their thoughts on the continuing need for the ASVAB.

Addressing the administrative components of the virtual meeting, Dr. Velgach said she needed a complete record of attendance and distributed an attendance sheet. She also informed participants that the meeting was *not* being recorded on the Microsoft Teams® system. She instructed all Teams participants to mute their devices and to click the "raise hand" button when they wanted to speak.

Dr. Velgach concluded her introductory remarks by thanking the committee members for their participation and the presenters for their support of the committee's activities. She then directed introductions of all participants.

The attendee list and agenda are provided in **Tab A** and **Tab B**, respectively. **Tab C** contains a list of acronyms. The Committee Chair has provided a letter, written by the committee members, summarizing key committee findings. The letter is included in these minutes at **Tab D**.

[1] The DACMPT website Meetings page is located at https://dacmpt.com/meetings/.

1. Accession Policy Introduction (Tab E)

Dr. Sofiya Velgach, Assistant Director, AP, presented the briefing.

Dr. Velgach began by summarizing the mission of AP, which is to "develop, review, and analyze policies, resources, and plans for Services' enlisted recruiting and officer commissioning programs." She then presented organization charts detailing the structure and programs within the Office of Under Secretary of Defense, Personnel & Readiness, as well as Deputy Assistant Secretary of Defense for Military Personnel Policy (DASD-MPP), and AP. An additional chart summarized some of the critical items facing the Directorate in various areas including recruiting, personnel security, testing, and medical qualifications. A number of recruiting challenges were cited, including the lingering effects of the Coronavirus (COVID-19) pandemic, minimal support from influencers of youth in recommending military service, low propensity of youth to serve, and the limited pool of qualified youth. Some mitigating factors are a professional and dedicated recruiting force, national support for a strong military, and robust virtual and social media engagement efforts.

A chart displayed data showing the level of propensity to serve in the military among youth ages 16-21, by gender, from April of 2001 through the fall of 2021. The latest numbers indicate that 9% of youth stated that it was likely they would be serving in the military in the next few years, including 11% of males and 8% of females. These low numbers make it harder for the Department of Defense (DoD) to make its mission. Another graphic indicated that, among youth 17- to 24-years old, 77% are disqualified without a waiver. Reasons include being overweight, drug abuse, medical issues, etc. Overall, 44% of those persons ineligible to serve are disqualified for multiple reasons. Among the 23% who are qualified without a waiver, 11% are enrolled in college.

A table displayed the fiscal year (FY) 2022 recruiting mission by Service and component (Active Duty, Guard, and Reserve), and recruiting and accession data. For Active components, only the Army fell short of meeting mission, achieving 75% of the goal of 60,000 accessions. Among Reserve Components, only the Marine Corps Reserve met its goal, with the other branches falling short by 25-39%. Dr. Velgach noted that in attempting to meet their recruiting mission for FY 2022, the Services depleted their Delayed Entry Pools (DEP), which set-up FY 2023 to be a challenging year. An additional chart showed that all components were able to meet recruit quality benchmarks in regard to education and Armed Forces Qualification Test (AFQT) category. The FY2023 starting recruiting mission includes 244,095 accessions across Services, along with approximately 29,000 officers.

Dr. Velgach concluded the briefing by highlighting high visibility testing efforts, including expansion of the ASVAB to alternate devices, development of joint enlistment and compatibility composites based on the Tailored Adaptive Personality Assessment System (TAPAS), and development of a complex reasoning test. Congressional requirements include examining ways to improve the military accession process and developing a computational thinking test relevant to military applicants.

As Dr. Velgach was describing activities associated with military naturalizations (slide 6), a committee member asked if individuals need to be US citizens – naturalized or other – to enlist. Dr. Velgach said they do not; lawful permanent residents (LPRs) are eligible to enlist, but non-citizens are unable to serve in sensitive positions (e.g., on a nuclear submarine). She said an advantage of serving as an LPR is access to an expedited naturalization process, potentially beginning on the first day of service. The committee member then called attention to the fact that the status of some non-citizens who are serving in the military has been reported to the media. Dr. Velgach said she would need to learn more about the situations and process before commenting further.

During the overview of high visibility testing efforts (slide 13), Dr. Pommerich added that each Service uses its own version of the TAPAS, but that the Defense Testing and Assessment Center (DTAC) is attempting to define a standardized version.

At the end of the briefing, a committee member asked for clarification on disqualifications related to weight. Dr. Velgach said the military measures and makes decisions based on body fat percentage (rather than weight) in accordance with a Department of Defense Instruction, but individual Services can be more restrictive.

A committee member, referring to general military propensity by gender (slide 8), asked if the recent increase in female propensity was viewed as a trend or a blip. Dr. Helland said it was a blip, adding that, as far back as the 1980s, female propensity has been driven by African American females. Two committee members then asked how to interpret the disqualification rates shown on slide 9, to which Dr. Velgach replied that the 77% disqualification rate is not indicative of the rate at which people are disqualified by AFQT scores. She said a score as low as 10 qualifies a person to serve, at least on that factor, and that the disqualification rate based on AFQT scores alone is only around one percent. A committee member asked if the one percent disqualification rate is so low because those who think they will fail do not apply. Dr. Helland said she believed that is a factor but those disqualified by the AFQT are often disqualified by other factors as well.

After Dr. Helland explained that the eligibility model shown on slide 9 was based on PAY97 data (i.e., the general market versus the applicant pool), Sergeant Major (SGM) Alan Myers (Headquarters Department of the Army, G1) commented that the Services can decide whether to disclose reasons for disqualification. He added that the Army has more flexibility in accessing low AFQT individuals, and that the other Services may – legally, in accordance with Title X – access up to 20% of applicants with low scores. Commander (CDR) Mike Pyne (Navy) asked if expanded testing would allow the Services to be more relaxed in regard to AFQT scores. Dr. Velgach said that was a good question, and that AP is trying to widen the pool of qualified applicants through use of the TAPAS. She explained that certain personality characteristics may help individuals overcome military educational challenges. She said AP is focusing on how to expand the force through expanded testing.

2. <u>New Member Briefing</u> (Tab F)

Dr. Mary Pommerich, Director, DTAC presented the briefing.

> Dr. Pommerich began the presentation by outlining the roles and responsibilities of various organizations associated with military personnel testing. These include: (a) DTAC, which provides research, development, and computer-based delivery of the ASVAB and special purpose tests; (b) the United States Military Entrance Processing Command (MEPCOM), which conducts testing operations and applicant processing; (c) AP, which maintains oversight of standards and applicable testing and joint-service oversight through the Manpower Accession Policy Working Group (MAPWG); and (d) the DACMPT, a federal advisory committee that provides independent, technical review and recommendations through AP to the Under Secretary of Defense for Personnel and Readiness (OSD P&R). She then presented organization charts for OSD P&R, DASD-MPP, and AP. Additional charts presented the organizational structures of the Defense Human Resources Activity, the Defense Personnel Analytics Center, and DTAC.

Dr. Pommerich continued by summarizing the principal projects under DTAC's purview, including the Enlistment Testing Program (ETP), the Career Exploration Program (CEP), the in-Service testing program, special purpose tests, and internet-based test delivery and psychometric consultation for DoD Language testing. DTAC's ASVAB mission includes (a) developing test items and forms; (b) calibrating, scaling, and equating new forms and item pools; (c) developing and maintaining test delivery software; (d) developing and maintaining a cloud-based test delivery platform; (e) developing and maintaining materials for the CEP; (f) monitoring operational testing; and (g) maintaining, analyzing, and disseminating ASVAB testing information. Research roles include evaluating ASVAB changes, conducting research on ASVAB use, evaluating and enhancing the CEP, and developing and evaluating new measurement methodologies.

Dr. Pommerich then provided an overview of DoD testing programs. Through the ETP, the ASVAB is administered to approximately 300,000 applicants annually at 65 Military Entrance Processing Stations (MEPS), two remote proctoring stations, and 178 Military Enlistment Testing (MET) sites. The Computerized Adaptive Testing ASVAB (CAT-ASVAB) is administered to almost all applicants at MEPS and MET sites, although paper-and-pencil (P&P) testing still occurs at remote sites with insufficient internet capabilities. Applicants can also take an unproctored version of the ASVAB, known as the Pending Internet Computerized Adaptive Test (P*i*CAT), followed by a short verification test (VTest) at the MEPS. Individuals who pass the VTest can enlist based on their unproctored scores; those who do not pass the VTest automatically take a proctored, full-length ASVAB. The APT is a short, unproctored screening test used by recruiters to identify applicants who are likely to qualify, however, APT results cannot be used for enlistment purposes.

Through the CEP, the ASVAB is administered to over 700,000 students in more than 13,000 high schools and post-secondary schools for career counseling purposes and the generation of recruiting leads. Currently, approximately 90% of all CEP participants take the P&P-ASVAB. Retired forms of the ASVAB are repurposed as the Armed Forces Classification Test (AFCT) which is administered to already enlisted personnel in order to determine academic readiness to pursue a different military career or enter Officer Candidate School. The program is administered by the Services, and most examinees take the web-delivered version of the AFCT, although P&P forms are available to sites without internet access.

Dr. Pommerich continued by showing slides detailing ASVAB test content and CAT-ASVAB and P&P-ASVAB test lengths and times. She then explained that the AFQT is a composite score based on verbal and math tests to determine enlistment eligibility. The AFQT is made up of a double weighted verbal composite (VE) (formed from an optimally weighted composite of unrounded Word Knowledge [WK] and Paragraph Comprehension [PC] scores) plus scores on Arithmetic Reasoning (AR) and Mathematics Knowledge (MK). All test scores are added as standard scores that are standardized on the 1997 American youth population. AFQT scores are reported on a percentile metric that represents the percentage of applicants scoring at or below the given composite score. "High quality" is defined as scoring above the 50th percentile on the AFQT. The Services have composites used to qualify applicants for specific military occupations. Dr. Pommerich then presented examples of Marine Corps composites.

Dr. Pommerich then turned to special purpose tests administered in the ETP. She noted that these are given at the request of the Services to determine qualification of applicants for specific occupational specialties or enlistment programs, and, while DTAC is responsible for the development and maintenance of the ASVAB, the Services and language proponents are responsible for special purpose tests. DTAC does provide delivery and consultation support. A table showed the tests currently in use or under development, including the Coding Speed test, the Cyber Test, the Defense Language Aptitude Battery, the TAPAS, Mental Counters, and Complex Reasoning. Another table showed which Service is using each test and identified the Service that is the proponent for each.

Dr. Pommerich then summarized the major DTAC research and development efforts. New items for the ASVAB are developed under contract, with DTAC editorial staff and subject matter experts (SMEs) providing oversight to ensure items are developed in accordance with standards and that item content is accurate and free of biased or sensitive content that could evoke adverse reactions from examinees. DTAC is also working with the contractor to develop and populate an item bank. There are four unique CAT-ASVAB items pools, with an additional pool used for the P*i*CAT. Historically there have been four P&P-

ASVAB forms for proctored administrations. They were created with four unique forms of the AFQT and two unique forms of the other tests in the battery which were scrambled to create four total forms. The current P&P-ASVAB forms, 27A and 27B were introduced following an incident of test compromise. Although the development of P&P-ASVAB forms has largely been discontinued because of resource constraints and their susceptibility to compromise, one last wave of four forms is under development for ETP use, planned for continuity of operations.

Four P&P-ASVAB forms are used for proctored administrations of all subtests in the CEP, with two unique forms of all subtests scrambled to create four total forms (23A, 23B, 24A, 24B). Forms 03E and 10E are used for proctored CAT-ASVAB administrations. For reasons stated earlier, P&P-ASVAB form development for the CEP has also been discontinued, with one last wave of four new forms under development for use in the CEP. The goal is to replace all operational P&P-ASVAB forms with CAT-ASVAB testing in the CEP for sites with sufficient bandwidth.

Dr. Pommerich continued by explaining that each examinee taking the CAT-ASVAB in the ETP receives 15 tryout or seeded items in two to four ASVAB subtests, with the total time spent answering tryout items being about 20 minutes. Examinees are randomly assigned to a tryout group that determines the subtests that will include tryout items, and these are randomly seeded throughout the tests. Responses to tryout items are not scored. After target sample sizes are obtained, seed items are calibrated and scaled to be on the CAT-ASVAB scale and analyses are conducted to screen items. Form assembly methods are used to build item pools from retained items. Scores on new pools are equated to scores on existing pools to ensure they can be treated interchangeably. Development of new forms/pools has historically been targeted at a 5-year cycle for both the ETP and CEP, although DTAC has had mixed success in meeting these goals. DTAC is working with the contract team to transition to developing new CAT-ASVAB forms every 2 years. An equating study is currently underway for CAT-ASVAB forms 11-15.

Dr. Pommerich then summarized other research and analysis efforts.

- Plan for the next generation of military personnel testing by evaluating the current ASVAB, obtaining input from stakeholders through focus groups, and conducting a survey of training developers and instructors to assess the relevance of the current ASVAB subtests to recruit training.
- Explore the potential for automatic item generation, focusing on Assembling Objects (AO), WK, MK, AR, and General Science (GS).
- Conduct research on the potential for using social media in recruiting.
- Conduct simulations to identify more efficient calibration methods.
- Evaluate the structure underlying the current CAT-ASVAB.
- Examine the status of the current norms and determine the need for renorming the ASVAB.
- Explore the potential for a Complex Reasoning test by developing an item generator, evaluating item performance, and developing a computational thinking composite.
- Evaluate adverse impact (AI) and differential prediction.
- Develop and expand upon validity arguments for the ASVAB, AFQT, and TAPAS.
- Explore the efficacy of machine learning methods to streamline ASVAB form development.
- Explore the feasibility of using Bayesian-based item calibration methods to reduce calibration sample size requirements.
- Conduct a high school curriculum study to assess the alignment between ASVAB content and current high school course offerings.

Dr. Pommerich continued by explaining that, historically, the CAT-ASVAB was administered at MEPS/MET sites via a local area network platform called WinCAT. In 2008, internet delivery was introduced on the Defense Manpower Data Center (DMDC) web platform. In 2021, DTAC transitioned ASVAB testing operations to an Amazon Web Services cloud platform. An internet version of the CAT-ASVAB (*i*CAT) is now solely administered via the cloud platform, as are most special tests. DTAC provides 24/7 helpdesk support, and data transfers are automated between DTAC and MEPCOM. The cloud platform has led to greater reliability and stability of test sessions, and previous platforms have been decommissioned. DTAC is in the planning stages to expand the devices on which the CAT-ASVAB can

be taken. Currently desktop, laptop, and Chromebooks are allowed, and recommended system specifications are provided. Plans are to expand testing applications to enable the use of mobile devices in some settings. This is based on research that showed no notable performance differences when examinees are familiar with the device used and the testing delivery interface is designed to be responsive to various device types. Device expansion programming efforts for *i*CAT are expected to commence in 2023.

Dr. Pommerich then turned to the CEP, which is a comprehensive career planning resource designed to help young people align their strengths and interests with post-secondary pursuits. It is offered at no cost and with no commitment to military service. Students in grades 10-12 and first/second year post-secondary schools are encouraged to participate, and those in 11th grade and beyond can use their CEP scores for enlistment. DoD sponsors the CEP with a two-part mission: Provide a free career exploration resource and provide qualified leads to the Military Services. MEPCOM administers the program and DTAC is responsible for technical development, maintenance, and evaluation. ASVAB was first introduced in high schools in 1968 and DTAC was assigned research and development responsibility in 1989. In 1992 when ASVAB 18/19 was implemented, significant improvements to the CEP were introduced, including the addition of the Find Your Interests (FYI) inventory. In 2005 the career development and planning model underlying the CEP was revised by establishing competency-based linkages between the three career exploration scores (verbal skills, math skills, science and technical skills) and the skill importance ratings in O*NET occupations. Two websites were also created (www.asvabprogram.com, www.careersinthemilitary.com) to provide students and counselors with career exploration and planning tools, including the OCCU-Find, a career database designed for high school students. Dr. Pommerich then showed a slide listing major developments in the CEP since 2014, which included redesigned websites, expanded military-civilian career crosswalks, post-test interpretation (PTI) training for those responsible for this function, and virtual program offerings to accommodate pandemic restrictions.

Dr. Pommerich continued by providing an overview of how the CEP works. Students take the ASVAB which helps predict future academic and occupational success. They then join a PTI session which includes an explanation of ASVAB scores, instructions for taking the FYI to identify their interests, and guidance in using this information in conjunction with the website and other career planning tools to help teens and young adults consider their options and make post-secondary plans that work for them and their family. Dr. Pommerich then showed a chart with participation numbers from 2019 to 2022, including the number of leads provided to the Military Services. In 2022, 607,920 students in 12,948 schools took part in the CEP which generated 497,164 leads to the Services. Dr. Pommerich then concluded by listing other CEP products, including ConCEP, which is a platform designed to consolidate access to ASVAB CEP applications and modernize processes and systems surrounding scheduling, marketing, and communications. Other efforts include establishing linkages between military and civilian occupations and U.S. Office of Personnel Management General Schedule job series, revamping the CEP website, and standardizing training for national field personnel.

Dr. Pommerich concluded her presentation by pointing out that the Services have traditionally taken the lead in research and development of non-cognitive assessments, but DTAC was recently charged with implementing pre-accession personality assessments for evaluation purposes. The goal is to expand enlistment eligibility and respond to recommendations of the Independent Review Commission on Sexual Assault in the Military. TAPAS has been identified as the assessment to be used for these efforts, and the current focus is on developing a Joint-Service TAPAS, examining the potential for a compatibility/suitability composite, and maximizing testing efficiency.

As Dr. Pommerich explained the organizational structure on slide 8, Dr. Velgach emphasized the extensive level of coordination that occurs between Accession Policy, DTAC, Joint Advertising Market Research and Studies (JAMRS), and other Office of People Analytics (OPA) directorates.

On slide 17, a committee member asked for clarification on the purpose of the AFCT. Dr. Pommerich explained that its content does not differ from that of the ASVAB, but that it is taken by personnel who have already enlisted. CDR Pyne explained that the Services use the test to qualify individuals for the more technical ratings and military occupational specialties (MOSs). On slide 21, another committee member asked Dr. Pommerich to explain the AFQT scoring process. Dr. Pommerich said the scores are standardized to a mean of 50 and standard deviation of 10. She also explained that equating is performed at the form level, which results in each form having a unique transformation to the mean and standard deviation.

After Dr. Velgach commented on the importance of the CEP PTI component (slide 43), a committee member asked who provides the interpretations. Dr. Velgach said PTIs are often given by the US Military Entrance Processing Command (MEPCOM) Educational Services Specialists (ESSs) who, along with recruiters, offer extremely valuable interactions with students. She said all personnel who provide PTIs are certified by the CEP. Dr. Jan Bayer (Written, LLC) added that school counselors are trained to provide PTIs as well.

At the end of the briefing, a committee member asked how AFQT verbal ability scoring accounts for English learners. Dr. Velgach referred to AP's policy statement on the subject: Military training is conducted in English; therefore, trainees should be sufficiently proficient in English to succeed in training and, subsequently, the field. She mentioned the Army's Foreign Language Recruiting Initiative (FLRI) Program, which identifies individuals who can sufficiently benefit from English proficiency training. In the FLRI, non-native English speakers must take a language proficiency test and a fluid intelligence test along with the AFQT. Those who qualify attend English proficiency training before starting basic military training. Dr. Velgach also said people often ask if the ASVAB is delivered in Spanish (or other languages) along with English. She said DoD is very attentive to this issue but needs to make sure recruits can pass initial training. She mentioned other issues with testing in multiple languages, involving psychometrics, costs, and decisions regarding where to place boundaries (i.e., which languages should be offered). She added that DTAC was looking for input from the committee on the matter, at which point a committee member asked if the other Services had similar programs. Dr. Velgach said the Air Force had conducted a trial program. Another committee member asked about item composition (i.e., are all items on the ASVAB subtests multiple-choice and do the special tests include many items with exhibits [graphs and charts]). Dr. Pommerich said all the subtests and special tests are multiple choice and some include figures and charts. She said the Coding Speed test is multiple-choice but uses a different type of question than the other tests. Dr. Pommerich also said Mental Counters is a dynamic test, which makes it different, and the TAPAS is forced choice with yet a different response format.

A committee member remarked that it would be interesting to learn more about the item types and their development, including sensitivity reviews and currency maintenance. Dr. Pommerich stressed that obsolescence is a concern, which is why DTAC wants a 2-year form/pool replacement cycle. Dr. Pommerich said DTAC could do a deeper dive on item development at the next DACMPT meeting. Another committee member said s/he would ask more questions about scaling later and thanked Dr. Pommerich for the detailed information. Dr. Velgach then gave a very brief overview of the documents that are available online.

The last topic of discussion addressed whether P*i*CAT administrations increased during the COVID-19 timeframe. A USMEPCOM representative indicated there is an increase in the use of P*i*CAT. Furthermore, Dr. Pommerich explained that due to COVID, the time authorized to complete the VTest after completion of P*i*CAT was extended from 30 days to 45 days. A committee member asked what percentage of people who take the P*i*CAT at home and require verification testing come in for the VTest. Dr. Velgach said the percentage is very high, because the people who take the P*i*CAT want to enlist; that is, they are not taking the test for practice. Exact numbers would need to be verified by MEPCOM.

3. <u>Milestones and Project Schedules</u> – (Tab G)

Dr. Mary Pommerich, Director, DTAC, presented the briefing.

Dr. Pommerich began the presentation with an overview of the projects to be covered in the briefing, including ASVAB development, ASVAB and ETP revision, the CEP, the Military Compatibility Assessment, and expanding test availability.

- New CAT-ASVAB Item Pools. The objective of this project is to develop CAT-ASVAB item pools 11 – 15 from new items. New item pool implementation is projected for summer 2023.
- Developing New P&P Forms. The objective of this project is to develop P&P ASVAB forms 29F/G, 30F/G, 31F/G, and 32F/G from new items. The projected completion date is to be determined.
- Evaluations of CAT-ASVAB and form development methodologies. The objective of this effort is to evaluate CAT-ASVAB methodologies and ways to streamline form development efforts. The projected completion date is to be determined.
- Next generation ASVAB/ASVAB analyses. The objective of this effort is to evaluate the state of the ASVAB and prepare for the next generation ASVAB and special purpose tests to be administered on the ASVAB platform in the ETP. The project is ongoing.
- Evaluating new cognitive tests: Complex Reasoning. The objective is to develop a non-verbal reasoning assessment and evaluate it for possible inclusion in the ASVAB and to develop an item generator for the Complex Reasoning test. Projected completion data is to be determined.
- Evaluating new cognitive tests: Computational Thinking. The objective is to develop a Computational Thinking composite score to meet National Defense Authorization Act (NDAA) requirements to address computational thinking skills. The projected completion date is to be determined.
- Evaluating new cognitive tests: Cyber Test. The objective is to develop a CAT version of the Cyber Test and program it for administration on DTAC's cloud platform. The projected completion date is to be determined.
- Evaluating new cognitive tests: Mental Counters. The objective is to refine the Mental Counters test of working memory and program it for administration on DTAC's cloud platform. The projected completion date is to be determined.
- Adding non-cognitive measures for selection and/or classification: TAPAS. The objective is to evaluate the use of TAPAS in military selection and classification. The project is ongoing.
- CEP. The objective is to revise/maintain all CEP materials (e.g., websites, print materials), conduct program evaluation studies, and conduct research studies as needed. The project is ongoing.
- Military Compatibility Assessment. The objective is to inform future development of an evidence-based accessions instrument. The project is ongoing.

- Expanding Test Availability: Web/Cloud Delivery of ASVAB and Special Tests. The objective of this effort is to program ASVAB and special tests for delivery on DTAC's web-based/cloud-based platform and introduce enhancements. The project is ongoing.

As Dr. Pommerich described the objectives and subtasks associated with the Computational Thinking test evaluation (slide 14), Dr. Velgach noted that the requirement for such a test was included in two successive NDAAs. A committee member asked how DTAC was defining "computational thinking," and Dr. Pommerich said there would be a presentation on that in a subsequent briefing. She said the NDAAs had given it a certain level of definition as a starting point.

In regard to the Military Compatibility Assessment (slide 24), Dr. Velgach said a separate Federal Advisory Committee – one looking at approaches to reduce sexual assault in the military – is tracking development of the assessment.

At the end of the briefing, a committee member asked in what ways the Next Generation ASVAB would be tied to the content of current high school courses, to include what is taught and how it is taught. Dr. Pommerich said the Next Generation ASVAB High School Curriculum Project was not far enough along to be briefed at this time, but that one or more HumRRO researchers might be able to comment on the matter. Dr. Tia Fechter (DTAC) responded that the initial plan is to look at course content and how it is taught, and how those things have changed over time. She cited examples including calculator use in math and how math is taught. The committee member reiterated the importance of staying current, to which Dr. Fechter responded that the last such comprehensive evaluation was conducted by Dr. Oppler over 20 years ago. She said the study found that individuals entering the Service had access to different types of courses than those who did not enter a Service, and, as such, it would be important to look at the current situations of those propensed to join the military versus the general population. The committee member responded that s/he thought DTAC was on the right track.

4. ASVAB Form Development (Tab H)

Dr. Matthew Trippe, HumRRO, presented the briefing.

Dr. Trippe began his presentation by presenting a graphic showing the steps involved in developing new CAT-ASVAB item pools. These include (a) administering tryout items to applicants, (b) calibrating and scaling the items, (c) conducting item screening (i.e., psychometric quality, differential item functioning), (d) identifying item enemies, (e) consolidating the analysis results, (f) using a simulation-based optimization algorithm to form item pools, (g) applying additional CAT parameters (e.g., exposure control, penalty parameters), and (h) equating pools. This was followed by tables showing the annual pool development targets and the current tryout item seeding design. The targets originally called for developing four pools per year for AR, MK, and PC, given that they are AFQT tests and have a moderate threat of compromise. Because WK is also an AFQT test and is the most susceptible to compromise, the plan was to develop eight WK pools per year. Because it has a moderate degree of compromise threat, four pools were to be developed for GS as well. Two pools were to be developed for each of the remaining tests.

Dr. Trippe went on to explain that tryout items are developed in "series" of 100 items per test. The convention has been that 200 items are required to develop one CAT pool. The tryout series are administered in "seed" versions, with Auto Information (AI), Assembling Objects (AO), Electronics Information (EI), Shop Information (SI), and Mechanical Comprehension (MC) requiring two series of 200 items; AR, GS, MK, and PC having four series of 400 items; and WK having eight series of 800 items. The

original annual pool development targets are proving to be too aggressive to support due to data collection requirements and demands on the psychometric and information technology teams. The item development and seeding design is being revised to be compatible with a flat target of 4-5 CAT pools every two years. CAT pools 5-9 will be operational from 2008 to 2022 and will be replaced by pools 11-15 which will be operational from 2023-2025.

CAT-ASVAB is based on the Three-Parameter Logistic Model (3PL). DTAC simulation studies of the calibration process suggested an item-level sample size of at least 1,000 is desirable for optimal parameter recovery of seeded items. Therefore, the target item-level sample size is 1,200 to account for some data loss associated with data cleaning (e.g., removal of corrupt or invalid records). Achieving the target depends on variable testing volumes but generally requires around 8 months of data collection. Each test is calibrated separately using BILOG-MG. DTAC simulations find that parameter recovery is improved as the number of seed items administered to each examinee increases. Parameter recovery is found to be relatively poor when 10 or fewer seed items are administered, so each examinee responds to 15 randomly administered tryout items per test according to seed design. Tryout items are calibrated in seed versions with 200, 400, or 800 items per calibration. This results in a sparse response data matrix with approximately 16,000 examinees for AI, AO, EI, SI, MC; 32,000 examinees for AR, GS, MK, PC; and 64,000 examinees for WK. To accomplish item parameter rescaling, the seed parameter values are calculated using seed response data with the latent distribution of theta fixed to BILOG defaults (0,1). Operational responses from the calibration sample are used in conjunction with the operational parameter values to estimate the latent distribution of theta on the operational scale for the calibration sample. Transformation constants are computed to put the seed parameters on the operational scale.

The psychometric quality analyses performed on each item include (a) reviewing item information, (b) examining item-model fit using eight fit indices, (c) content review of distractor analysis, (d) conducting differential item functioning analyses, and (e) applying a screening rubric to identify items ready for operational use, items ineligible for operational use, and items that require further review to determine eligibility. Dr. Trippe then displayed a chart showing the percentage of items per subtest that were retained during the development of Forms 11-15.

Dr. Trippe continued by discussing item enemy analyses. In 2008, Drs. Pommerich and Segall evaluated local dependence (LD) in CAT tests and found that LD in item parameters has minimal effect on precision, while LD in item responses has a substantial effect. Mitigating LD requires identifying enemy groups of items likely to trigger LD if they are administered to the same person because they measure similar or highly related content. Before assembling forms 5-9, DTAC developed a content framework for identifying enemy groups for tests where LD is of particular concern. These are MC which covers 111 content areas and MK with 212 content areas. CAT-ASVAB ensures an applicant is administered no more than one item from an enemy group. For the other tests, there is no empirical evidence of LD affecting item responses, but it is also true that some items assess similar content. Evaluating the degree of content similarity among a matrix of greater than 1,000 items is challenging, but HumRRO has developed methods to optimize human labor and machine learning and natural language processing roles. Dr. Trippe then detailed the process used for WK and the process used for the other ASVAB tests (minus AO).

Dr. Trippe continued by explaining that CAT administration is based on pools from which a potentially unique set of items is administered to each examinee. Because of this, the pools need to contain items from the full range of content and difficulty and sufficient information/score precision across the full range of ability. The pool assembly goals include (a) assigning one item to one of five item pools for each test, (b) maximizing the precision levels of each pool, (c) constraining the precision levels to be comparable across pools, (d) accounting for enemy items by distributing them evenly across pools, and (e) accounting for content taxonomies where applicable (e.g., GS, AO). Dr. Trippe then presented an example of how the process works in developing a WK item pool and charts showing the score information for (a) pools 11-15 and (b) pools 11-15 and 5-9 compared to P&P. He continued by explaining how the Sympson-Hetter exposure control algorithm is applied to reach an overall item exposure rate of 1/6 in the ETP.

Dr. Trippe then turned to a discussion of penalty parameters. CAT-ASVAB scores, which are derived from the Bayes Modal Estimator approach, contain bias that draws the estimate toward the mean of the prior.

This type of bias tends to be larger in shorter tests like CAT-ASVAB. Low-ability examinees could potentially exploit this by answering the minimum number of questions allowed. A simulation-based penalty procedure is employed that assigns a final score equivalent to the expected score obtained by random guessing on the unanswered questions. The penalty functions are regression equations.

Pool equating is implemented in three phases of operational administration of new pools to military applicants, with each phase including progressively larger sample sizes. The intent is to maximize the accuracy of the reported operational scores. A random groups design is used with each applicant assigned to a single pool with a 1 in 7 assignment probability. Reference Form 4 is administered only during equating studies along with an operational form and a new form (11-15). The qualification composite cumulative distribution functions (CDFs) between reference Form 4 and the new pools are evaluated. Dr. Trippe then presented charts showing the qualification rate differences between (a) the reference form and the new form and (b) the operational form and the reference form.

Dr. Trippe continued by explaining that CAT-ASVAB pools 11-15 were administered to applicants in May 2022 as part of the equating study, the first two phases of which are complete. The phase 3 sample size is projected to be achieved by mid-December 2022. A modern computing workflow was developed for CAT-ASVAB pools 16-20 which will run in parallel with the original Fortran-based processes. Since the assembly of pools 11-15, 28 WK series have been processed, as have 12 series for the AFQT tests and GS, and 4 series for the technical tests. Next steps include completing phase 3 equating for pools 11-15 by computing final transformation constants and implementing a thorough evaluation and analysis. Development of pools 16-20 will begin using eligible items from the new series processed since assembling 11-15, items not assigned to the 11-15 pools, and items not assigned to P&P ASVAB forms.

Dr. Trippe then turned to a discussion of P&P form development by showing a graphic outlining the steps involved. These include (a) identifying eligible items, (b) scaling the items, (c) checking for adherence to content specifications, (d) identifying enemy items, (e) examining target form characteristics (i.e., test information functions, reliability), (f) examining target population characteristics (i.e., latent distribution, composite conditional distributions), (g) implementing automated test assembly processes, and (h) conducting an SME review of the results. The goal is to develop new P&P ASVAB forms to replace those currently in use in the CEP and ETP. The CEP has four forms (23A/B, 24A/B), where A and B versions include the same items reordered. The ETP also has four forms (25A/B, 26/B) with the A and B versions containing unique items for the AFQT tests and the same items reordered for the non-AFQT tests. This effort will likely be the last wave of P&P form development. Eligible items for the P&P forms include those developed for CAT-ASVAB pools 5-9 and 11-15 which were not assigned, and eligible items from item series processed since development of CAT pools 11-15. The P&P and CAT-ASVAB are on separate scales. DTAC previously conducted an anchoring study to link the P&P-ASVAB and CAT-ASVAB scales. Latent means and standard deviations from that study were used to apply linking constants in reverse to place item parameters scaled to CAT-ASVAB on the P&P-ASVAB scale.

Each P&P-ASVAB test has a content blueprint specifying the number of items and the sub-content distribution (e.g., whole number and rational numbers for AR; life science, physical science for GS). The CEP and ETP test blueprints are the same. Dr. Trippe showed a table detailing, for each test, the number of items and time limits for CAT-ASVAB and P&P ASVAB. An Automated Test Assembly optimization model is used to develop forms parallel to each other and "target" CEP/ETP forms. Model constraints include number of items, content blueprint, item key balance, item enemies, maximization of test information functions (TIFs), and minimization of equally weighted sum of the distance between TIFs and test characteristic curves of the forms. Quantitative evaluation criteria include the similarity to "target" CEP/ETP form TIF, alignment with latent distributions, and alignment with latent distributions conditional on aptitude area composites. A final SME review is conducted to identify item enemies, obsolete content, and sensitive content. Dr. Trippe then showed examples of test information plots for WK and GS.

Dr. Trippe continued by discussing some of the technical challenges faced in P&P form development. These include the fact that when the P&P-ASVAB was originally developed, AI and SI were calibrated and scaled as one test. For CAT-ASVAB, items are calibrated and scaled as separate tests and then combined into a composite. P&P-ASVAB must include Auto & Shop (AS)-scaled item parameters to be compatible

with MEPCOM infrastructure. Rescaling options include conducting a special data collection to administer new AI and SI items along with backup and reserve items and original AS items, then conducting calibration and final form assembly. This is impractical and risky. Another solution is to apply a modified Stocking-Lord procedure (MSLP) for two tests/scales transformed to a common scale. This involves iteratively trying out sets of transformation constants (A [scale] and B [location] constants) and searching for the set that best minimizes the objective function. The objective function is the sum of the squared differences between (a) the expected number-correct scores based on sets of parameters on the AI and SI scales and a simulated distribution of true-score AI and SI thetas and (b) the expected number-correct scores based on a set of parameters that have been rescaled using provisional constraints and the average of the true-score AI and SI thetas. Two simulations were conducted to compare MSLP to item parameters estimated by calibrating AI and SI items together in BILOG-MG. The results of both were consistent and showed a very close correspondence between TCCs for MSLP- and BILOG-scaled parameters. There were slightly lower TIFs for MSLP, but the BILOG-based TIFs are likely inflated due to violating the assumption of unidimensionality.

Turning to PC, Dr. Trippe explained that when the P&P-ASVAB was originally developed there were five questions per paragraph stimulus. CAT-ASVAB items are developed with one question per stimulus. Maintaining a fifteen-item PC test would result in increasing the number of paragraphs from three to fifteen. The addition of twelve paragraph stimuli would dramatically increase the word count and thus the time limit. Testing time is valuable and increasing the time limits would be problematic for both the CEP and ETP. The solution involves modifying the automated test assembly optimization algorithm to include existing constraints (e.g., content blueprint, item key balance) and including a new constraint to minimize the projected response time. The variable of interest is the number of items (9-15). Findings suggested that the ten-item solution is optimal for minimizing response time while minimizing loss of test-level reliability and maintaining composite reliability. The current P&P ASVAB PC time limit is 13 minutes, which is a test session time limit applied to all test takers. DTAC will provide a recommended solution at the upcoming Military Accession Policy Working Group (MAPWG) meeting when the details are finalized. Dr. Trippe then showed charts indicating there is an obvious and strong relationship between the number of items and word count, which leads to longer testing times. With a 15-item solution the projected time limits would be greater than 20 minutes, nearly doubling the current 13-minute limit, which is unacceptable. Additional charts showed there is a weaker relationship between the number of items and the reliability of PC scores and the reliability of AFQT percentile scores. The ten-item solution represents the optimal compromise between R_{xx} and projected testing time. DTAC will prepare a recommendation based on comprehensive research.

Dr. Trippe then summarized the current status and next steps in P&P form development. All solutions are completed except for PC for both the CEP and ETP. They have been quality control checked and are ready for delivery to DTAC. The final AS solutions were mostly unaffected by the scaling decision, but some "what if" scenarios regarding the order of operations are still being evaluated. The ETP MK solution is ready for delivery, but another "what if" analysis is underway as PC and AS solutions are being fully resolved and implemented. Next steps include finalizing the PC solution under a unified approach and developing six parallel PC forms (2 CEP, 4 ETP) using the latest optimization algorithm. The AS scaling decision needs to be made, but the MSLP approach seems preferable to additional data collection. The research findings and resulting recommendations will be presented to the MAPWG. Then all P&P form deliverables will be finalized, and attention will be returned to CAT-ASVAB pool development. Dr. Trippe concluded by soliciting DAC feedback on various issues, including the AS and PC technical challenges.

As Dr. Trippe began the briefing, Dr. Velgach emphasized the highly technical nature of the slide content and reminded everyone that discussion time is limited.

The first question dealt with WK score information (slide 15). The committee member asked whether the information function indicates that the goal of maximizing conditional precision levels of each new pool has been met and, if not, what are the implications for examinees and what could be done about it. Dr. Trippe said he interpreted the plots as showing that the forms do

well in maximizing score information where most people would be affected by the assessment. He said the evidence used to support this conclusion is the relationship between the score information functions and the distribution of theta. Dr. Pommerich clarified that they are not looking for a uniform distribution but are trying to maximize the information where there is the most density, as WK scores are used only in composites. The committee member agreed but asked about the implications on cut scores (i.e., where decisions are made). Dr. Trippe said that was an interesting question, though there are no cut scores on the individual ASVAB subtests. The committee member asked if one could evaluate the theta estimates for a subset of examinees who are borderline qualified to get a sense of the range for each subtest. Dr. Trippe said they had computed latent distributions conditional on aptitude area composites and identifying where decisions are made is the best way to identify the most important parts of the distribution. He also said he could provide an example on slide 27, which showed the conditional latent distribution plot for the P&P-ASVAB forms. He explained that the template for the presentation should be adjusted in the future because the committee member's question had been asked in previous meetings.

Another committee member asked two questions, the first being on the relationship of local dependence and enemy items, and the second on the topic of item compromise. That is, what is the rationale for judging certain tests (e.g., Shop) as being under lower threat of compromise (on slide 5)? Is it because once people learn the knowledge, they know it, while conversely, people can train for a math test without learning the underlying knowledge and skills? Dr. Trippe agreed with the committee member's explanation on item compromise and explained that there are different ways content can be compromised. He said harvesting occurs when people try to memorize content and distribute it on the Internet. He explained that WK items are very short and easy to memorize, but MC content is more difficult to memorize and harder to harvest. Dr. Pommerich said another issue is that the AFQT tests have higher stakes, which is part of the reason they are on the "high compromise" list. She then explained how the matter was moot to a large degree, because new forms for those tests are now targeted to be introduced every 2 years. She said they had initially wanted to replace content for WK every 6 months and for the other AFQT tests every year, but that was too ambitious. Returning to the committee member's first question on local dependence and enemy items, Dr. Trippe said they identify the items that are similar enough to produce local dependence and split those items among pools. He said, if there are more enemy items than pools, then the items are evenly distributed among pools. Finally, he described an additional parameter in the CAT engine that continues to identify enemy items as such, so that any given applicant will not see more than one.

On qualification rate differences (slide 19), a committee member said it sounded like the calibration of seed items was performed on the same operational scale across forms (i.e., Services). S/he asked why it is necessary to use linear transformation to put everything on a reference scale. Dr. Trippe said that was a good question and that the committee member understood what was occurring. Dr. Trippe said he had characterized the action as being "potentially unnecessary" and explained that other testing programs do not do it and consider the rescaling step sufficient for equating. He described the extra step as insurance. Dr. Trippe explained that, because the work was not finished, he did not have a plot that shows the extra step is doing something, even if it is not a lot. He reiterated that its implementation is meant to provide additional confidence that replacing pools is not disrupting the continuity of the testing

program. The committee member asked, however, if the procedure might be introducing bias and, therefore, doing more harm than good. Dr. Trippe said there will be a stand-alone briefing on equating and that prior analyses (and the associated briefing) had explored the possible introduction of bias.

At this point, Dr. Velgach advised that the rest of the briefing should focus on P&P-ASVAB tests, including the specific technical challenges and questions DTAC had for the committee.

At the end of the presentation, a committee member asked for a brief description of the passages that are used in the PC items, noting that the length and depth of passages affect what can be measured. Dr. Trippe explained that the original P&P passages were deep and long but that the newer passages are approximately three to five sentences in length. He said they are taken from open-source texts whose copyrights have expired and are deep enough to support the associated questions. Dr. Pommerich said the length of the new passages is intended to eliminate the need for scrolling on the CAT-ASVAB forms.

Dr. Trippe asked for the committee's advice on dealing with the challenges faced on the AS and PC subtests. Dr. Velgach asked if the committee had recommendations on the approach or concerns with the methodology. A committee member responded by asking if different means and standard deviations necessarily indicated multidimensionality. Dr. Trippe said it was a conundrum and the ideal solution would be that used for the CAT-ASVAB form. He said they know AI and SI are distinct, though they have some overlap, but in CAT-ASVAB, they are scaled separately and scored as a composite. He said using that approach would be preferred, but it is prohibited until they have some MEPCOM infrastructure changes for the P&P-ASVAB form, leaving the current scaling approach as the best approach for the P&P-ASVAB form. The committee member suggested DTAC might not find a lot of differences regardless of which approach they implemented. Dr. Trippe said, as a practical matter, the alternate approach is quite painful due to requirements such as having to trick the system to administer the trial items to the same applicants in a randomized group design on the original AS forms. He said calibrating items together in such a manner would run the risk of violating unidimensionality assumptions. He said it seemed like a lot of effort for an approach that might ultimately fail. He credited Jeff Dalkhe (HumRRO) for his work on the task and offered to talk in more detail with the committee member in coming weeks.

5. Next Generation ASVAB/Testing – Evaluation Plan (Tab I)

Dr. Mary Pommerich, Director, DTAC, presented the briefing.

> Dr. Pommerich began by stating that the purpose of the briefing was to provide background and an update on the history, status, and plans for the next generation of the ASVAB and special tests administered on the ASVAB platform as part of the ETP. This will include an overview and status of current efforts followed by more details on several of these efforts, including the ASVAB and AFQT validity frameworks, the training relevance survey, and focus groups conducted with ASVAB stakeholders.
>
> In 2005-2006 the ASVAB underwent a systematic review with testing experts making recommendations for improvements and enhancements to the ETP. The panel was motivated by a difficult recruiting environment and the belief held by some that the ASVAB was outdated and in need of an overhaul. The MAPWG condensed and prioritized the panel's recommendations. Dr. Pommerich then showed a list of those

recommendations and noted that all of them have been investigated and evaluated, with changes implemented where warranted. Some research and development efforts are still underway (i.e., Complex Reasoning test). Although many changes have been introduced in the ETP since the review, the content of the ASVAB itself has not changed. Prior discussions on possible changes have stalled due to (a) a lack of consensus about the philosophy of the ASVAB, (b) logistical challenges associated with making changes such as dropping tests that would have an impact on current composites and the systems set up to operate on those composites, and (c) concerns about insufficient resources to accommodate a revised ASVAB that would require more testing time than the current battery. However, in 2022 there is once again a challenging recruiting environment, and some hold the belief that the ASVAB is outdated and in need of an overhaul. DTAC hopes to respond to questions via Next Generation Testing efforts. The key steps included are:

- Develop and/or refine and evaluate special tests of interest.
- Evaluate the existing ASVAB tests.
- Collect focus group information and use that to develop a shared vision for Next Generation Testing.
- Consolidate and synthesize the information gathered and use that to make recommendations about individual tests.
- Form and convene an ASVAB Stakeholder Advisory Committee to help guide decision making about changes to introduce in Next Generation Testing.

Dr. Pommerich continued by outlining some key objectives for a revised ASVAB/ETP, including (a) improving selection into the military and classification into MOS; (b) maintaining testing standards and current value while introducing improvements or enhancements where needed and practical; and (c) introducing changes that meet diverse stakeholder needs which will likely involve compromises, as those needs vary widely.

Next Generation Testing efforts are focused on the ASVAB as well as special tests that are administered along with the ASVAB in the ETP on the ASVAB platform. The special test scores are used in addition to ASVAB scores for classification purposes. Due to limited testing time, it is necessary to simultaneously consider all tests to be administered. Also, any changes to the ETP must be take into consideration the impact on the CEP.

Dr. Pommerich continued by providing a progress report on these efforts.

- Efforts are ongoing to develop and refine special tests of interest such as the TAPAS, Cyber Test, Mental Counters, and Complex Reasoning.
- Evaluation of tests currently in the ASVAB is also ongoing, as are efforts to apply an argument-based approach to validation of the ASVAB and AFQT.
- Focus groups with stakeholders and documentation of the results is in progress.
- An effort has been initiated to revisit the question of the philosophy of the ASVAB as a first step in developing a shared vision that defines the purpose and general makeup of the ASVAB and ETP for Next Generation Testing.
- Consideration is being given to an appropriate methodological approach to synthesize the evaluations of the ASVAB and special tests and make recommendations regarding Next Generation Testing.
- Forming an ASVAB Stakeholder Advisory Committee is a potential future effort.

Dr. Pommerich then provided an overview of the status of special tests of interest. Improvements in instructions and practice items are being introduced to the Mental Counters test to eliminate a persistent floor effect in the applicant population. A CAT version of the Cyber Test will be introduced to better target item difficulty to applicant ability. The Complex Reasoning test, which assesses fluid intelligence, is under development modeled after Raven's Progressive Matrices. Efforts are underway to develop a joint-Service version of the TAPAS and to use TAPAS for the assessment of military compatibility/suitability and to maximize test efficacy.

Total testing time will be a key consideration if there is a desire to incorporate additional tests in the ASVAB given resource constraints that limit the ability to accommodate increases in testing time. A comprehensive assessment of the tests currently in the battery will provide insight into their utility, quality, and potential for

modification. Potential changes to accommodate new tests include dropping and/or combining and/or shortening existing tests or merging new tests with existing tests. DTAC has an extensive effort underway to evaluate the current ASVAB tests to determine their desirability or expendability. Dr. Pommerich showed a list of these efforts, including evaluating item/form development costs, examining test content durability, and evaluating the efficiency of each test along with its susceptibility to compromise. In formulating ASVAB evaluation summaries, the goals are to identify a way to concisely summarize the results across all steps and a way to aggregate the findings and compute an overall rating for each test. Dr. Pommerich then described several approaches DTAC has considered. She concluded her presentation by seeking DAC guidance on ways to (a) synthesize the qualitative and quantitative data that are being gathered, (b) move forward with identifying possible changes if there is no consensus on the philosophy of the ASVAB, and (c) move forward in the face of widely varying stakeholder needs and interests.

As Dr. Pommerich described the key objectives (slide 7), a committee member asked how DTAC might define improvements in selection (e.g., increases in validity or job satisfaction). Dr. Pommerich said answering that question would likely require another look at the philosophy or purpose of the ASVAB. She said this is the very question DTAC has been trying to answer and that it is being addressed – in part – in ongoing discussions with stakeholder groups regarding the potential need for compromise in revising the ASVAB.

At the end of the briefing, Dr. Pommerich asked for the committee's thoughts on several key issues. Dr. Velgach reiterated that any continuing discussions between DTAC and the committee would need to be held publicly, perhaps at the next DACMPT meeting. Dr. Pommerich responded that it would be better to have feedback sooner than later, because HumRRO would be moving forward on the development of the philosophy of the ASVAB, and a task order for working toward a synthesis approach would be starting in June 2023.

A committee member asked how DTAC had learned that the needs of the stakeholders vary so greatly. Dr. Pommerich said the information had been collected through focus groups and Dr. Adams (HumRRO) would brief on that, to include how focus group members were selected. Another committee member noted the value, but also impossibility, of having a single purpose for the ASVAB. S/he asked if meeting the stakeholders' goals is a mutually exclusive proposition or if there is a way to meet everyone's goals. Dr. Pommerich said a completely shared vision is likely impossible and compromises will likely be needed. She said, for instance, military trainers are very pleased with the current tests because they identify people who can succeed in training. Alternatively, she said recruiters want a test that will bring in more people, allowing them to meet their recruiting missions. She also described how there is more agreement among stakeholders on the selection component than on the classification component. In all, she reported being encouraged by what she heard from the focus groups, to include the fact that there were no major complaints. Dr. Pommerich ended the exchange by noting that a primary issue lies in the area of public relations, especially with respect to questions about why test content has not changed a lot over time. She added, however, that the program is anything but static and that DTAC is working on information papers and other communications to better educate users.

A committee member said that reducing test length is a common theme in the testing world. S/he said tailoring for different stakeholders is one way to increase efficiency, but the requirement to integrate across branches complicates the situation. S/he added that shortening testing times may be better accomplished by using a multi-dimensional approach, where one concept informs

another concept (e.g., multi-dimensional item response theory [IRT]). The committee member said a taxonomy content review might also identify redundancies.

The committee member then asked what goals in relation to continued ASVAB changes might be pulled from consideration. Dr. Pommerich said switching to different information technology systems would be a major undertaking, and the current software development lifecycle process is very cumbersome due to cyber security requirements, which make it difficult to enact mission changes.

The committee member also raised the possibility of shorting test administration times by changing CAT stopping rules and asked if DTAC could use existing data to change the rules. Dr. Pommerich said DTAC could look at that, but the tests are already short. She also said many tests are not content balanced, which prohibits shortening those tests. Dr. Pommerich then referred to MEPCOM's desire to process recruits in one day without having them stay overnight, which precludes additional testing. Dr. Pommerich also mentioned item seeding requirements, noting that DTAC might need to move seeding efforts from proctored ASVAB administrations to VTest administrations if P*i*CAT usage continues to increase. A committee member mentioned that testing time is always an issue in private industry, though studies have shown that test-takers are tolerant of lengthy tests, as long as they view the material as job-relevant. S/he recommended that DTAC consider what test lengths recruits are willing to tolerate. Dr. Pommerich said she had not heard complaints about test times from applicants, but that the complaints about test length come from other stakeholders. Dr. Velgach advised DTAC to consider collecting more information to further assess the situation.

A committee member asked Dr. Adams to elaborate on her statement that trainers were generally satisfied with the test. Dr. Adams explained that trainers teach new jobs to recruits and believe the training covers the aptitudes required for successful training. Alternatively, she said recruiters have to encourage people to take the ASVAB because there is already resistance to testing. The committee member said, on the recruiting side, the issue approximates one of validity due to the number of persons who choose not to apply because they dislike testing. S/he asked Dr. Adams to comment on how potential applicants view the process and if they think they are prepared. Dr. Adams replied that her team had talked with 20 applicants and recruits but had not been able to access high school students taking the CEP because focus group members had to be 18 years or older and required permission from their school. Dr. Adams also said her team has not captured perceptions of the ASVAB from applicants who have completed training but asked Dr. Velgach if that could be an option in the future. Dr. Velgach explained that similar information (i.e., how do you feel your score or experience correlates to your actual ability level) is captured on the New Recruit Survey, and results indicate the test is well aligned with ability self-assessments. A committee member asked if range restriction might be in play, given that only applicants who were accepted completed the survey. Dr. Pommerich responded that it would be interesting to talk with applicants who were not accepted and summarized by emphasizing the importance of understanding the environment.

6. Next Generation ASVAB/Testing – ASVAB/AFQT Validity Framework (Tab J)

Dr. Deidre Knapp, HumRRO, presented the briefing.

Dr. Knapp began by explaining that the purpose of the validity argument frameworks is to compile, organize, and review existing evidence related to the use of pre-enlistment assessments. In this context "relevant information" is defined much more broadly than psychometric properties or criterion-related validity evidence and includes all aspects of a measure's design, development, administration, score reporting, and so forth. The process seeks to evaluate whether the available evidence supports the use of the assessments for their intended purposes, identify ways to strengthen that evidence, and help inform improvements to the assessments in terms of content, scoring, administration, and/or interpretation.

Dr. Knapp continued by stating that a theory of action (TOA) is a useful starting point for developing a validity argument. The TOA consists of a set of connected propositions that explain a specific goal of the assessment and, as such, TOAs are intimately linked to the purposes and uses of assessment scores. The interpretive argument is a description of the inferences that the assessment scores are intended to support (i.e., claims, assumptions). The validity argument consists of evidence providing justification for the inferences in the interpretative argument. Dr. Knapp then showed a chart summarizing these components.

The first effort focused on use of the AFQT for enlisted selection. A second effort focused on the use of TAPAS for enlisted selection and classification. Development of a validity argument for ASVAB use for classification decision-making is in progress, as is a revision and update of the AFQT validity argument. Dr. Knapp noted that the nature of this work is dynamic as validity argument evidence evolves over time. She then showed a chart presenting a portion of the TOA for the AFQT and another displaying the interpretive argument structure. The AFQT interpretive argument is being revised to reflect an improved approach and format adopted for the ASVAB Classification interpretive argument. This will have minimal impact on the types of evidence collected but will ensure that the identified assumptions for the ASVAB and AFQT interpretive arguments are the same or largely the same to simplify future updates to both. Dr. Knapp then showed an evidence summary example for the first version of the AFQT validity argument. The format includes a description of the assessment and the validity argument approach. There is a table for each major claim that lists the associated specific claims and assumptions with hyperlinks between the assumption and the applicable evidence summary. A narrative summary of the evidence for each major claim is also provided and recommendations are included. For the AFQT, these included the philosophy underlying test content decisions, suggestions for future research, and administrative improvements.

Products associated with this work include (a) recommendations to help make decisions about future research and development efforts, (b) an organized archive of nearly 500 references (e.g., technical reports, briefings, journal articles), (c) dozens of succinct research summaries on various topics, (d) an organizing framework into which new evidence can be placed and communicated, and (e) a detailed process map and supporting tools for periodically updating each validity argument. Dr. Knapp concluded by noting that it is challenging to develop clear, concise, and comprehensive interpretative arguments. The primary audience for evidence summaries needs to be considered in their development, and those summaries benefit from many layers of review. She also noted that no attempt was made to quantify the evaluation of the evidence and that probably is not warranted.

At the end of Dr. Knapp's presentation, a committee member asked, "Who is the primary audience for the work?" Dr. Knapp said the main audience is researchers who work for the various Services. The committee member commented that the effort is quite impressive and provides a useful encapsulation of a tremendous amount of information about the ASVAB. S/he said the task, which s/he then briefly summarized, probably would have been a lot easier for a test that was just being developed. Dr. Knapp responded that the team had tried not to hone in too much on criterion-related validity but rather conduct a broader evaluation of the test from administration to the reporting of scores and everything in-between.

Another committee member said a similar effort had been attempted on another national standardized test; however, that effort had been approached more as hypothesis testing and turned out to be extremely difficult as well as less successful. The committee member suggested one question should be whether the correlation between measures is lower at higher ability levels and/or whether classification power might be more useful at higher levels. S/he commented on how questions like these could guide the building of a TOA. Dr. Velgach commented that senior military leaders are asking questions about the value of standardized testing. Is there value in terms of selection, classification (ability to identify the best fits between different aptitude levels and jobs), and/or development programs? She asked the committee for their thoughts. In response, one committee member raised the possibility of leveraging the power of noncognitive tests (e.g., motivation) to evaluate applicants in a more holistic fashion and identify compensatory abilities. S/he also described how tests can be used in a less deterministic manner and used more to discover areas in which applicants need help or improvement (e.g., math). Another committee member stated that standardized testing provides a means for objective evaluations and a way to make high staked decisions.

Another committee member said s/he really liked the framework, which s/he said was much better than the old approach because it allows an examination of all aspects of the test. The committee member also said that validity is not constant or enduring and asked if communication to primary audiences should be considered as a validation claim. Another committee member reflected on the importance of "prioritizing the audience" and said the work was very good.

7. Next Generation ASVAB/Testing – Training Relevance Survey (Tab K)

Dr. Scott Oppler, HumRRO, presented the briefing.

Dr. Oppler began by explaining that the purpose of the training relevance survey was to evaluate the relevance of ASVAB content, as well as other test content/constructs, to success in entry-level military training. The approach involved administering a survey to technical training course developers and instructors to collect ratings of relevance and content knowledge/skill to success in technical training for entry-level military occupations. This replicated and extended work done in the late 1990s, which was limited to content associated with the four ASVAB science and technical subtests. In this extension, the content areas covered all of the ASVAB subtests as well as the Cyber Test and a hypothetical test of computational thinking. SMEs were asked to rate relevance on a four-point scale:

- This knowledge is not relevant for the training.
- Trainees entering training with this knowledge/skill perform no better than those entering without it.
- Trainees entering training with this knowledge/skill perform somewhat better than those entering without it.
- Trainees entering training with this knowledge/skill perform much better than those without it.

They were asked to make ratings based on three knowledge levels:

- Basic—basic knowledge of facts, definitions, or terminology.
- Application—skill to follow procedures or use different types of tools, equipment, or mathematical equations associated with specific technical content areas.
- Analysis—skill in using reasoning and analytic abilities, including the skill to troubleshoot a situation, fault isolate a symptom, or solve problems associated with specific technical areas.

Dr. Oppler then showed an example of the format used to collect the ratings. This was followed by an overview of the courses included in the sample. There were 160 courses across all six Services covering a range of technical areas (e.g., electronics, mechanical, information technology, intelligence). For each science and technical subtest along with AO and the Cyber Test, two databases were created. One included courses using the test as part of their classification composite and the other for courses not using the test. For each dataset for each test, the data were analyzed within course and (a) across all courses, (b) by training focus area, and (c) by Service. For each content area within each test, the percentage of courses in each of the four relevancy categories was calculated. A single dataset was created for the AFQT tests, quantitative reasoning (QR), verbal ability (VA), and computational thinking. For each construct domain, the data were analyzed (a) within course and across course, (b) by training focus area, and (c) by Service. Again, the percentage of courses in each of the four relevancy categories was computed. Dr. Oppler then displayed a sample table of results.

Dr. Oppler continued by providing a summary of the results for each subtest along with tables displaying those results.

- AS. For training courses with AS in their selection composites, relevancy ratings generally support the AS content categories, although somewhat more so for AI content and "Shop Tools." Results varied across training focus areas in expected ways (e.g., relevance ratings for Mechanical courses were generally higher than for Electronics courses).
  - AS content was generally rated to be much more relevant for those training courses with AS in their classification composites than for those without it.
- EI. For training courses with EI in their selection composites, relevancy ratings generally support all EI content categories. Like AS, ratings for EI generally varied across training focus areas in expected ways (e.g., relevance ratings for Electronics courses were generally higher than Mechanical courses).
  - EI content was generally rated to be much more relevant for courses with EI in their composites than for those without it.
- GS. For training courses with GS in their composites, ratings were generally low for most of the GS content categories with the exception of some of the Physical Sciences (Force/Motion, Energy, Fluids). Among the training focus areas, GS content categories were rated the highest (and for a greater variety of content categories) for Science and Engineering courses.
  - GS content was generally rated to be only slightly more relevant for courses with GS in their composites than for those without.
- MC. For training courses with MC in their composites, ratings generally supported most of the MC content areas with the possible exception of "Structural Support." Results generally varied across training focus areas in expected ways (e.g., relevance ratings for Mechanical courses were generally higher than Electronics courses).
  - MC content was generally rated to be much more relevant for those training courses with MC in their composites than for those without.
- AO. For training courses with AO in their composites (all of which were Navy), the relevancy ratings for the one AO content area (Spatial Ability) were generally high, although the sample size was relatively modest ($k = 17$).
  - Except for courses in the intelligence training focus area, the relevance ratings for AO were also relatively high for courses that do not include AO in their selection composites, suggesting that this test may be underutilized.
- AFQT QR Content. Content categories included in the survey under QR were intended to represent content on the MK and AR subtests. Results generally supported all QR content areas, although relevancy ratings were highest for Numbers and Operations and Measurement.
  - Relevance ratings varied across training focus areas to some extent, with ratings for Science and Engineering courses generally higher than those for other areas.
- AFQT VA content. Content categories included in the survey under VA were intended to represent content on the WK and PC subtests. Relevancy ratings generally supported all VA content categories and were greater than those for QR.
  - Ratings varied across training focus areas to some extent, with ratings for Intelligence courses generally higher than those for other areas.

> Dr. Oppler summarized the results by stating that they support the relevance to training performance of the vast majority of content categories included on the ASVAB. The primary exception to this may be GS, for which only a small number of content areas, primarily in the Physical Sciences, received ratings in support of their relevance to training performance. Results for the Science and Technical subtests were very consistent with those from the 1997 study. Results for AO suggest it may be underutilized given the relatively high relevancy ratings it received for training courses that do not currently use AO for classification. Results for the Cyber Test suggest that it also may be underutilized, at least for courses in some training focus areas (e.g., Information Technology and Intelligence). Results for Computational Thinking suggest that this construct may be relevant to training performance across a wide range of training focus areas, although ratings were highest for courses in Information Technology and Intelligence. All six Computational Thinking content areas received support, with the highest ratings for Problem Decomposition, Pattern Recognition, and Abstraction. Dr. Oppler concluded by soliciting feedback from the DAC on how DTAC can use these results as they consider the Next Generation ASVAB as well as any other comments they may have.

When Dr. Oppler explained that the AO test may be underutilized (slide 31), Dr. Velgach said only the Navy is using it currently, though the Air Force has done much work in modeling and is evaluating whether it can incorporate the test.

In response to Dr. Oppler's solicitation of questions and feedback at the end of the briefing, a committee member asked for clarification on the rating process. Dr. Oppler said they averaged the ratings of SMEs for each course. He added that if a course was used by multiple Services, the ratings were averaged across Services for those Services that used a particular subtest in their classification composite for that course.

A committee member raised a potential issue with the utilization of the Cyber Test, asking, "Does the test inadvertently screen out potentially qualified candidates?" S/he said not all high schools offer computer science courses covering the topics addressed by the Cyber Test. S/he added that the classes, if offered, are expensive, which discourages participation by those on the lower end of the income distribution. Dr. Oppler said the same could be said for some of the other technical tests. He then referred to AS, explaining that prior research had found students who joined the military were more likely to have taken an auto or shop class. He said he was not sure how this applied to the Cyber Test but suggested that many people learn computer skills outside of school; that is, if people are interested in computers, they find ways to learn about them. He said the Cyber Test may operate as a maximal measure of interest. Dr. Sophie Romay (Air Force) commented that the Cyber Test had increased, rather than reduced, the candidate pool for certain jobs and that it has served as a counterbalance against the AI imposed by the Mechanical Comprehension and Electronics Information tests. Dr. Velgach asked the Navy representatives if they use the Cyber Test, and CDR Pyne said they do. He said he had heard people refer to it as being akin to a measure of interest. Returning to the original concern, Dr. Velgach reiterated that, even if someone is interested, there still may be barriers to learning. Dr. Pommerich said Dr. Greg Manley (DTAC) would discuss the small to moderate effect sizes that he has observed in relation to the Cyber Test. Dr. Kimberly Adams (HumRRO) noted the Cyber Test's concentration on topics like security, networking, and communications.

Another committee member asked if range restriction might have been an issue with the ratings. S/he also asked if tests may not have been associated correctly with courses due to issues related to rater style or tendency to think a course was not relevant because of thinking such as, 'I can

teach anything.' Dr. Oppler reported his confidence in the ratings, largely because of the similarity between the aggregated results from the current and prior studies. He also said the results make a lot of sense; endorsements of courses were in the direction that one would expect. He added that raters were told not to focus on the tests they use to select personnel into their MOS or ratings. Dr. Adams clarified that the SMEs were not asked to link tests to courses but just to rate them. She said SMEs from the Manpower Accession Policy Working Group (MAPWG) were asked to help make connections between the courses and tests.

8. Next Generation ASVAB/Testing – Focus Groups (Tab L)

Dr. Kimberly Adams, HumRRO, presented the briefing.

Dr. Adams began by explaining that the purpose of this study was to collect information from a variety of stakeholders about their perspectives on the ASVAB and the ETP. Questions focused on such topics as what tests should be administered as part of the ASVAB or on the ASVAB platform in the future, and what other changes are needed to modernize the ASVAB and ETP. The potential stakeholder groups included MAPWG members, individuals from the military/DoD realm (e.g., recruiters, trainers, policymakers), representatives from the educational arena (e.g., MEPCOM ESS), members of national and state boards of education, high school teachers and counselors), and applicants, recruits, high school students and their influencers. Dr. Adams continued by showing a chart that identified the subtests included on the ASVAB from 1968 to the present. This was followed by another chart showing special tests currently administered on the ASVAB platform as part of the ETP and the Services that use each test.

Dr. Adams continued by showing the list of stakeholders annotated to indicate which groups have already taken part in the study, including those from the military/DoD realm, MEPCOM ESS, career counseling organization, high school counselors, and applicants/recruits. Each focus group included a representative sample of 10-12 participants. Those from the military realm were familiar with the ASVAB and its relationship to their role. Education representatives were also familiar with the ASVAB and the CEP and were from geographically dispersed areas covering the entire country. Examinees were 18-years or older and included both applicants and recruits who had taken the ASVAB in the ETP. They were also representative in terms of Service branch they were entering or applying to enter and in regard to geographic region. Each focus group was virtual and lasted 2 hours. The sessions were recorded and transcribed, but participants were assured that the results would be treated confidentially, and individuals would not be identified in reporting. They were instructed to speak from their perspective only and told that the goal was to gather information, not achieve a consensus. Non-government participants were provided a small honorarium for taking part. Dr. Adams continued by identifying recruiting sources for each group and listing the number of groups per stakeholder category along with the number of participants.

For the military/DoD stakeholders, the focus group questions centered on (a) what participants like and dislike about the current ASVAB, (b) what changes to the ASVAB/ETP they would make today, (c) primary reasons for changing the ASVAB/ETP in the future, (d) barriers to such changes and how they can be overcome, (e) goals to be obtained with a revised ASVAB/ETP, and (f) the priority that should be given to diversity and inclusion in considering revisions to the ASVAB/ETP. The notes from each focus group were combined, and coding schemes were developed using whiteboards generated during their discussions. A qualitative data analysis tool (MAXQDA) was used to code the notes, and consensus meetings were held to discuss coding differences. An inclusion rule was implemented that required three or more stakeholder groups to comment on a category for it to be included. Quality and consistency reviews were conducted for all coded data.

Dr. Adams continued by presenting a summary of the military/DoD stakeholder group findings. The top three positives regarding the ASVAB were that it predicts performance success, improves person-job fit, and measures a wide range of knowledge and skills. In each case, these pluses were mentioned in 16 of the 35 military/DoD focus groups. The top three negative aspects of the ASVAB were that it does not measure

potential/ability to learn (16 of 35 groups), it includes content that is outdated or inapplicable to the Services (15 of 35 groups), and it does not predict all aspects of performance/success (e.g., adaptability; 13 of 35 groups).

Dr. Adams concluded by presenting the next steps in the project, which included completing the analyses of the military/DoD results, conducting analyses for the Education and Examinee groups, and documenting the findings both within and across groups.

At the end of the briefing, Dr. Adams requested guidance and solicited questions from the committee. She specifically asked for feedback on the number of unique focus groups (i.e., three needing to comment on a theme in order for the theme to be included in the results). A committee member responded by asking if that approach reduced the possibility that input by participants who took the opposite side of an issue might not be captured. Dr. Adams said their recording and analyses accounted for disagreement on topics (e.g., focus group participants were asked what they "like" about the ASVAB/ETP as well as what they "dislike" about it). She also said a review was conducted of all comments made by only one focus group to ensure no single statement offered a unique insight that would elevate for its inclusion in the final report. Also, DTAC will receive a data file of all de-identified comments to review independently. The comments were coded separately for the three different realms of stakeholders (i.e., DoD/military, education, and examinee stakeholders), updating the coding scheme developed with the DoD/military focus group data to capture the other stakeholder groups' perspectives. The report will include a separate section for each realm of stakeholders, given there were differences in the participant recruitment strategies, variations of the protocol questions, and coding schemes. She said the objective is to provide a report that synthesizes and summarizes the results in a manner that it can be used to understand what each realm of stakeholders likes and dislikes about the ASVAB/ETP, what they would like to change or not change, and why.

Another committee member asked if the participating examinees reflected the diversity in the Services. Dr. Adams said participants did not provide demographic information, but efforts were made to ensure the recruitment of focus group participants took that into consideration. For the DoD/military focus groups, she said focus groups were conducted separately for each of the six Services to support a diverse sample across military branches. She also worked with MAPWG Service representatives, Mr. Davis at MEPCOM, and their designated POCs to promote regional and occupational diversity among the participants within each stakeholder group. Based on introductory information, the participants provided about themselves at the start of the focus group sessions, as well as their responses to protocol questions, there appeared to be regional and occupational (when appropriate) diversity achieved.

A committee member said the number of participants in the study was incredible and that s/he was thinking of all the journals that would be interested in the study. The committee member specifically mentioned American Psychological Association Division 5. Dr. Velgach pointed out that there are restrictions on publishing DoD work. Another committee member then asked how the team handled the fact that qualitative text was interpreted by different coders. Dr. Adams said it took a year to collect the data due to the large number of focus groups and noted the dataset of comments is massive. She said they are looking at best practices in the literature to make sure they are using the most appropriate inclusion criteria (i.e., number of focus groups making a comment for it to be included in the results summary). She reiterated that they want to communicate everything that is important, even if did not meet reporting thresholds. She said, to

be part of the coding team, a person had to have been a notetaker in at least three focus groups. A committee member asked if MAXQDA transcribed accurately, but Dr. Adams said it was not set up for that task. Dr. Pommerich said they used teams of notetakers. Another committee member asked if the questions that were asked of military personnel were modified for non-military groups, and Dr. Adams said they were.

9. Norming Requirements/Plans (Tab M)

Ms. Pamela Baumer, HumRRO, presented the briefing.

Ms. Baumer began by presenting a chart summaring the development of the ASVAB and past norming efforts. The original norms were based on the World War II mobilization population. These were in place until 1980 when new norms were developed based on a study done in conjuction with the National Longitudinal Survey of Youth (NLSY) conducted by the Department of Labor (DoL). This was repeated in 1997, but for various reasons new norms were not put into place until 2004. Since that time, DTAC has conducted various evaluations of the norms by examining changes in performance on various national tests such as the National Assessment of Education Progress (NAEP) and Scholastic Aptitude Test (SAT), to identify any major shifts in abilities among the youth population. The results of these analyses have not suggested an urgent need for a new norming study. In 2022, a working panel will be convened to explore avenues for future renorming efforts and discussions are underway with the DoL regarding coordinating efforts in relation to the 2026 NLSY.

Ms. Baumer continued by stating that in 2017, the DAC provided guidance that it is not the age of the norms, but the validity of the norms and changes in the norming group that are most important to consider. In regard to validity, questions include "Does the norming group represent the target population? Are the norms based on a valid methodology? and Have changes in the populaion resulted in the ASVAB not measuring the same things as it did in 1997?" The norming group should be assessed to see if there are changes in abilities and/or demographics and determine if those changes have a meaningful effect on AFQT scores. Ms. Baumer continued by showing a chart displaying mean NAEP math and reading scale scores from 1980 to 2012. She noted that there have been moderate increases in both scores for Black and Hispanic youth, which yielded a small increase for the overall population in 1996. However changes from 1998 to 2012 were negligible. Another chart provided population demographics in 1980, 1997, and 2017. This indicated a growth in the Hispanic population, fairly steady percentages of Black youth, and a decline in the "other" category. An additional chart showed highest educational attainment of youth in these same years, which indicated a decrease in the percentage of high school dropouts and high school graduates and increases in those with some college or a college degree. Finally, Ms. Baumer showed a chart showing the percentage of youth scoring in various AFQT categories in 1980, 1997, and from 1997 to 2017. She noted that there were some neglible differences between 1980 and 1997 in regards to Categories IIIB, IVC, and V, with a maximum difference of 4 percent. However, there were no differences between the 1997 percentages and estimates since that time (timeline stops at 2017).

Ms. Baumer then discussed the steps for addressing the need for a new norming study, noting that DTAC would like to see an annual estimate of changes in ability and demographics along with an assessment of the impact of those changes on AFQT estimates. The steps involved forming a working group of experts, holding relevant discussions, and setting the criteria for when a new norming study should be conducted. The working group should include a demographer, psychometrician, sampling statistician, and someone with experience with large-scale norming studies. Their discussions should focus on (a) what data are most useful in estimating changes in youth abilities and the impact of those changes on ASVAB norms, (b) how estimates of ability change and their impact on ASVAB norms can be generated for subgroups for which there are no data, and (c) how the data can be clearly and concisely presented to decision makers. In setting the criteria, the primary question is what amount of change in abilities/demographics would be needed to justify the expense of conducting a new norming study, taking into consideration statistical significance, effect size, consistent multi-year changes, and historical criteria.

Ms. Baumer continued by noting that the 1997 norming study had a 77 percent response rate, which was lower than the 1980 response rate, and said Hispanics were particularly underrepresented in the sample. In addition, the weighted sample educational levels were dramatically higher than expected based on other available data. The Bureau of Labor Statistics (BLS) is developing a needs assessment for methods and research questions for the 2026 NLSY. An expert panel was included to explore how DoD could be a partner in this effort to obtain measures of cognition, personality, and career interests of youth. Similar arrangements to previous efforts are under discussion. Among the considerations are the strengths and weaknesses of various approaches, the costs involved, and relevant research studies that may inform planning. Ms. Baumer closed by asking the committee for their thoughts on several issues, including estimating ability changes for groups when there are no available data, defining meaningful change in AFQT scores that should trigger a new norming effort, other methods for collecting the needed norming data, and how to take into consideration the impact of COVID-19 on educational acheivement.

After Ms. Baumer explained the NAEP long term trend test results from 1978-2012 (slide 6), a committee member asked if she had a similar plot for AFQT scores. Ms. Baumer said DTAC should be able to provide one.

When Ms. Baumer described the need to determine which data would be most useful for estimating ability change and its impact on ASVAB norms over time (slide 11), a committee member asked if Ms. Baumer had access to ASVAB standard scores. Ms. Baumer said she did not, noting that the chart on slide 8 was based on percentile scores. She added that using standard deviations might be an option. The committee member then asked if applicants receive standard scores when they take the ASVAB. Dr. Pommerich responded that DTAC has a substantial amount of ASVAB data that would allow examination of how scores are changing over time. Dr. Velgach elaborated, saying AP can see in the current data that there are smaller proportions of very low and very high aptitude persons in the applicant pool. Dr. Pommerich explained that numbers are highly related to employment conditions; that is, good employment conditions result in fewer applicants, especially those of high aptitude. The committee member commented on the importance of examining trends given the frequency of job market shifts. Another committee member suggested that conducting measurements on a more frequent basis could allow for predicting or imputing measurements – though with different confidence intervals – for a given year or group as long as past data and the characteristics for the group are available.

Another committee member inquired as to whether Ms. Baumer's team is looking into score trends pre- and post-COVID. Ms. Baumer responded that they are and asked if the committee had any recommendations on how to approach the matter. A committee member referred to a previous effort in which one large national test was re-normed based on trends witnessed over a few years of data. S/he also reference a study by Dorans. At this point, the discussants referred to a trend in which high school test scores have been lower post-COVID and observed that no one knows if the trend will be temporary or enduring. They mentioned contributing factors, including virtual learning. Ms. Baumer asked for the committee's recommendation on proceeding with the re-norming in the current environment. One committee member said that s/he would not recommend an extensive re-norming examination until more time had passed. Dr. Pommerich agreed and stated that the cost of renorming is almost prohibitive, and the earliest effort would probably take place in 2026.

Ms. Baumer continued the discussion by mentioning that NAEP scores have declined in recent years. A committee member said that was not surprising and added that s/he had seen increased

subgroup differences as well, probably due to vulnerable populations being differentially affected by recent conditions. The committee member recommended that the team be sensitive to changes resulting from those types of factors over time.

When Ms. Baumer described the BLS's development of a needs assessment for methods and research questions for a 2026 NLSY (slide 15), Dr. Velgach called attention to BLS's interest in working with AP to identify areas in which the programs can be better aligned. Dr. Pommerich noted that there are many moving pieces in play and said the two organizations are trying to resolve timeline issues that may impact more extensive collaboration.

To conclude the briefing, Ms. Baumer asked for the committee's thoughts on several issues she had addressed. Dr. Velgach asked the committee if it would be appropriate to develop norms based on the applicant pool versus the entire population, though she noted that norming based on the population was customary. She also asked about the potential impact of increasing score validity versus obtaining a better understanding of the market. CDR Mike Pyne referred to the philosophy of the ASVAB and suggested that basing norms more strictly on the applicant pool would not be a problem because the test is used primarily to select and classify persons into the Services.

A committee member asked Ms. Baumer if her team had reviewed methodologies utilized in the education field. Ms. Baumer said she had not done an extensive literature review in that area. Dr. Pommerich commented that DTAC typically requires literature reviews as a first step. The committee then discussed whether national norms for the American College Test (ACT) were updated yearly, though they drew no conclusion. Dr. Velgach closed the discussion by referencing the methods employed by BLS. Some of the methodological issues being considered by BLS are the tests to include in the norming effort, proctoring methodology, and sample sizes.

10. <u>Device Expansion Plans</u> (Tab N)

Dr. Tia Fechter, DTAC, presented the briefing.

> Dr. Fechter began by explaining that the goal of this effort is to expand the number and variety of devices (e.g., tablets, smart phones), web browsers, and operating systems that can be used to take unproctored assessments such as P*i*CAT and APT. This will allow for more flexibility for ASVAB administration, reduce time spent in the MEPS, and increase the number of applicants. This will require the development of a Next Generation user interface that incorporates a Responsive Design approach that automatically formats the test to display on alternative devices.
>
> Dr. Fechter continued by reviewing the history of device evaluation research related to the ASVAB. A previous study found that the specific device an examinee uses to take the ASVAB does not significantly impact test scores. However, that same research found examinees perform better on some subtests of the ASVAB when they are familiar with the device they use as opposed to not being familiar with it. In general, examinees took less time in responding to items on alternative devices in comparison to the Dell XPS notebook; however, differences were not practically significant (less than 30 seconds for a full length ASVAB subtest). She noted that previous studies have shown that examinees are currently provided sufficient time for responding to test items on allowed administration devices. In regard to whether device type differentially impacts item difficulty, the answer appears to be most likely not. Of all 1,116 comparisons, 7.4 percent were flagged for DIF based on device type, where a Type I error rate of 5 percent suggests that many of these differences are likely to be unexplainable. In regard to whether item features (e.g., inclusion of a graphic) have an impact on item difficulty, no effect was found for those features studied. Two post-hoc reasons for DIF for

some items were classified into (a) preference for an item display for some items with certain features and (b) ease of looking up answers via an internet search. Another question examined was whether there were interactions of device familiarity and historically disadvantaged subgroups that impact score performance on ASVAB subtests. The one instance found that the positive effect of familiarity with device was heightened for the lower SES subgroup on MK; the score differences widened for those unfamiliar with the device they used to take the ASVAB. Dr. Fechter concluded that, overall, these findings suggest that ASVAB subtest scores among applicants should be comparable regardless of the device used to take the tests, so long as the examinee uses a device that is familiar, and the test delivery application is designed to be responsive to a variety of device types.

Dr. Fechter continued by discussing operational implementation considerations such as who should take the ASVAB, on which mobile device, and for what purpose. A wide variety of options were discussed, although many were acknowledged as not being practical at this time. The option could be extended to applicants taking the test at home or at a MEPS/MET site, as well as students participating in the CEP. Devices could be owned and maintained by the test taker, DoD, or schools. Options also include the unproctored APT, which is a screener and does not yield a score of record, the P*i*CAT, which requires a verification test for enlistment, or proctored sessions at MEPS/METS or in high schools. Among the factors to be considered are the potential for test compromise if examinees using their own devices take screenshots, maintenance costs of DoD-owned devices, and score effects associated with testing on unfamiliar devices.

Dr. Fechter then turned to recommendations and progress. The first recommendation is to design a test delivery application for the ASVAB that is responsive to a variety of device types. Operational implementation decisions must be made prior to moving forward with device expansion. The initial focus will be on devices commonly used by those eligible for military service or by students in high school. Smartphones will be limited to low-stakes testing (i.e., APT) and ASVAB tests for which scores must be verified (i.e., P*i*CAT) to mitigate security concerns. Browser and operating system options will also be expanded where feasible. Dr. Fechter then showed a chart illustrating a three-phase implementation plan for unproctored tests, with desktops and laptops using expanded browsers and operating systems being available in Phase 1, followed by tablets and touchscreen technology in Phase II, and Smartphones in Phase III. Examinees will be allowed to choose the device they are familiar with to take the ASVAB. Dr. Fechter noted that in the September 2020 DACMPT meeting, a committee member warned that research has suggested that students may lack the ability to make optimal decisions regarding essay prompt selection and was curious whether these findings would generalize to selecting a device for testing purposes. This will be mitigated by providing clear advice to examinees as to which devices and device specifications would be optimal for test taking.

Additional implementation steps include developing a test-monitoring plan that tracks operational performance differences (scores and response times) between device types. This was recently completed, including a literature review on unproctored mobile testing that resulted in additional recommendations to DTAC. Those recommendations included the development of (a) a data-collection tool that reports device features (e.g., screen size, browser type and version, device type) for post-test monitoring and analysis; and (b) a post-test questionnaire intended to measure barriers to optimal performance. Dr. Fechter then showed two slides listing the data currently captured during ASVAB administrations and the proposed additional data to be collected when device expansion is implemented. Data from tests administered on alternate devices will be compared to those on standard devices at the item, subtest, composite, and battery level. Among the variables of interest are the device/operating system/browser used and examinee familiarity with them, test administration conditions (e.g., distractions, technical issues), demographics, features or characteristics of items or tests, and others to be determined. The types of analyses to be conducted include various generalized linear models and differential item functioning.

The frequency of monitoring will depend on sample size conditions, and sample size targets will be developed prior to initiating each analysis taking into consideration statistical power, Type I error, and the expediency required to answer questions. Monitoring will be ongoing until there is sufficient evidence that device type does not significantly impact performance and will resume when device types change or are expanded. The monitoring will isolate triggers for performance differences (e.g., screen scrolling requirements) and will include remediation plans. Moving forward DTAC will seek advice from consultants with in-depth

knowledge of the practice of administering high-stakes assessments on mobile devices to ensure that all implications have been considered. Recruitment efforts in this regard are underway. When device expansion has been implemented, DTAC will systematically monitor the data on an ongoing basis and create a template for reporting outcomes to be used by decision makers to promptly address any necessary issues. Dr. Fechter concluded by soliciting the committee's input on other considerations that should be incorporated into the implementation plan or any other issues they may have.

When Dr. Fechter finished the presentation, a committee member commented on the importance of user interface and individual characteristics. S/he said familiarity is part of the issue, but there are likely interactions between familiarity and type of task and content format. S/he suggested collecting data on the nature of the task, to include how content is displayed, as well as the individual's knowledge of how to move around in that specific context. The committee member said there is interesting research in that area. Dr. Fechter said they had conducted a small usability study during the interface design phase, the purpose of which was to make the interface more intuitive. She agreed with the committee member that knowing how to use a device does not necessarily mean you know how to use it in testing. She said they specifically targeted an evaluation of AO because of Susan Embretson's earlier findings that placement of the response options in relation to the stem affected item difficulty. She said, in their study, they did not find this to be the case for the specific 1x4 and 2x2 layouts used. She also said Coding Speed and Mental Counters presented challenges, but longer passages in PC items did not appear to affect item difficulty, which she thought was surprising.

A committee member asked if DTAC had considered requiring a "lockdown browser" for use with the P*i*CAT. Dr. Pommerich explained that they cannot control what individuals put on their devices, but that they do use that approach at MEPS facilities. The committee member asked if the P*i*CAT was all multiple choice, and Dr. Fechter said it is, at least presently.

Another committee member complimented Dr. Fechter on the presentation. The committee member, commenting on slide 13, said that differential analyses are challenging, and the effects are small and difficult to interpret. S/he also said that looking at all the factors matters, especially because those with fewer resources are in more challenging environments. S/he expressed interest in knowing how DTAC will handle the data and deal with interaction effects, to include potentially discarding items or changing technologies. Dr. Fechter said all these things are why they want to talk to people who have done this type of work. The committee member then said that, hopefully, the data will come from the bottom up, and that a nice plan will allow for that. Dr. Pommerich said that is why they are starting with the APT and P*i*CAT, since they are low-stakes tests. P*i*CAT has a VTest component which decreases the risk of error. She said it would be a larger concern if they were attempting to use alternate devices at MEPS. Dr. Velgach said there is some interest in allowing alternate device use in schools, but that they want to start with unproctored and low-stakes versions.

The discussion ended with a committee member's comment on the potential for isolating effects in a single study, to include identifying interaction effects. S/he suggested that main effects, as well as interaction effects, may point to the need to adjust scores, though s/he noted the statistical and political issues that might arise. Dr. Fechter agreed and said that is why they had conducted both analysis of variance and Bayesian analyses, with the latter being able to look at a wider variety of interaction effects.

11. Public Comments

After the end of the first day of presentations, Dr. Velgach opened the floor to public comments and asked participants to limit their comments to no more than 5 minutes per person.

Mr. Davis commented on the time constraints at MEPS, saying the issue is that the bussing process is predicated on the administration of the ASVAB. The addition of the TAPAS in 2009, and presently, other special tests cause testing time to bleed into time required for medical processes. A committee member asked if recruits completed all the tests in a single day and whether they could return for a second day. Mr. Davis said they try to process people as quickly as possible. He also said they have a single day cycle (i.e., AM and PM), as well as a night-time cycle in which they test from 1300-1700 people who stay overnight and return for medical the next day. He mentioned that they try to get medical results while the recruits are on site. He said they have not been able to gain back time at any part of the process leaving them continually squeezed.

Dr. Velgach suggested that administering the Cyber Test only to those who have done well on certain other tests might be a way to save time. A committee member noted that time was not the only issue, but that there appear to be philosophical issues at play as well. Mr. Davis said they try to minimize the time recruits spend at MEPS so that everything, to include swearing-in, occurs during a one-day process.

Mr. Matthew Scheidt (USMC) made two points. First, it is important to understand a recruit's abilities and to know that they can be trained to standard. He said that includes knowing whether they might already have technical knowledge that reduces the amount of technical training they will need to complete once they enlist. Second, the Marines are anxious to see how the use of fluid intelligence tests might eliminate some of the adverse consequences of using math and verbal tests or more general intelligence tests. He said he is hoping that such a test might expand the boundaries of who can qualify. Mr. Davis said that MEPCOM works closely with the Marines, because they use the P*i*CAT so frequently, which shortens ASVAB testing time at the MEPS and allows more time for other tests and processes.

12. ASVAB Adverse Impact (Tab O)

Dr. Greg Manley, DTAC, presented the briefing.

> Dr. Manley began the briefing by defining AI as the unintended discrimination of a protected class that is the result of a selection procedure. AI is not a property of a test per se. However, AI may occur when a test's scores are used as the basis for selection. When a test is not directly used in selection (i.e., individual tests that collectively comprise the selection composite), the 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 score difference gives us an index to examine a test's potential for AI. The four-fifths rule is often used to determine the occurrence of AI: "A selection rate for any race, sex, or ethnic group which is less than four-fifths (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. Dr. Manley then presented a formula for calculating the impact ratio and for determining the significance of that value and the confidence intervals around it.

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 (a) examinees who qualify for entry into the military (i.e., those scoring in AFQT category IIIB or higher, AFQT percentile ≥31, the most common selection cut score used by the Services) and (b) examinees who qualify for enlistment incentives (i.e., those scoring in AFQT category IIIA or higher, AFQT percentile ≥50). Dr. Manley noted that the initial test scores were the only test score of record used for all AI analyses (i.e., scores from retests or confirmation tests were excluded from the analyses). Dr. Manley also noted that significance testing is not particularly useful for analyses with very large numbers of applicants (i.e., >2000). Effect sizes (i.e., standardized mean score differences) provide a method of evaluating potential for AI across individual ASVAB and Special Tests, where no direct selection occurs. Dr. Manley then presented a formula for computing effect sizes and their confidence intervals. He explained that effect sizes can be plotted and classified with respect to Cohen's standards of evaluation, with small effect sizes starting at 0.20, moderate at 0.50 and large at 0.80. He explained that the ASVAB testing program evaluates AI by comparing males and females, non-Hispanic Whites and Hispanic Whites, non-Hispanic Whites and non-Hispanic Blacks, and non-Hispanic Whites with non-Hispanic Asians. He noted that non-Hispanic Asians now represent more than 2% of the applicant population.

Ideally, AI is assessed on a regular basis. DTAC's longitudinal analysis program has examined AI for every odd-numbered fiscal year (FY) since FY2005, excluding FY2007. The data presented here were used to assess AI for applicants testing in FY2021 (October 1, 2020 through September 30, 2021). Dr. Manley noted that FY2021 includes parts of the COVID-19 "shutdown" year and the year following the shutdown. Additionally, there are issues with missing demographic data, which sometimes resulted in sample sizes for race/ethnicity that were very small and may not be representative of the overall population of actual applicants. However, the effect sizes are similar to previous years and the analyses should provide insight into the effort to remove barriers that negatively impact diversity. Dr. Manley then displayed a chart showing the FY21 sample sizes for each of the groups of interest. He continued by presenting charts showing the following results:

- Impact ratio and 95% confidence interval for FY 2021 AFQT cut scores IIIB and IIIA, all education levels
- Comparison on impact ratios for FY09 through FY21
- Comparison of effect sizes for odd numbered FYs 09-21, males versus females, AFQT Tests/Scores
- Comparison of effect sizes for odd numbered FYs 09-21, males versus females, non-AFQT Tests
- Comparison of effect sizes for odd-numbered FYs 09-21, non-Hispanic Whites versus Hispanic Whites, AFQT Tests/Scores
- Comparison of effect sizes for odd-numbered FYs 09-21, non-Hispanic Whites versus Hispanic Whites, non-AFQT Tests
- Comparison of effect sizes for odd-numbered FYs 09-21, non-Hispanic Whites versus non-Hispanic Blacks, AFQT Tests/Scores
- Comparison of effect sizes for odd-numbered FYs 09-21, non-Hispanic Whites versus non-Hispanic Blacks, non-AFQT Tests
- Comparison of effect sizes for odd-numbered FYs 13-21, non-Hispanic Whites versus non-Hispanic Asians, AFQT Tests/Scores
- Comparison of effect sizes for odd-numbered FYs 13-21, non-Hispanic Whites versus non-Hispanic Asians, non-AFQT Tests

Dr. Manley summarized the results by stating that the magnitude of impact of 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 FY2021 magnitudes are reasonable. Sufficient information is available online for two other large-scale testing programs: the SAT Math and Reading tests taken by 2016 college bound seniors and the 2019 NAEP grade 12 reading, math, and science tests. Dr. Manley then presented a series of charts showing the following:

- Comparison of effect sizes across testing programs, math content area, males versus females
- Comparison of effect sizes across testing programs, math content area, non-Hispanic Whites versus Hispanics
- Comparison of effect sizes across testing programs, math content area, non-Hispanic Whites versus non-Hispanic Blacks
- Comparison of effect sizes across testing programs, math content area, non-Hispanic Whites versus non-Hispanic Asians
- Comparison of effect sizes across testing programs, reading/verbal content area, males versus females
- Gender representation across samples/populations
- Comparison of effect sizes across testing programs, reading/verbal content area, non-Hispanic Whites versus Hispanics
- Comparison of effect sizes across testing programs, reading/verbal content area, non-Hispanic Whites versus non-Hispanic Blacks
- Comparison of effect sizes across testing programs, reading/verbal content area, non-Hispanic Whites versus non-Hispanic Asians
- Comparison of effect sizes across testing programs, science content area, males versus females
- Comparison of effect sizes across testing programs, science content area, non-Hispanic Whites versus Hispanics
- Comparison of effect sizes across testing programs, science content area, non-Hispanic Whites versus non-Hispanic Blacks
- Comparison of effect sizes across testing programs, science content area, non-Hispanic Whites versus non-Hispanic Asians

Dr. Manley continued by stating that 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 the impact of ASVAB tests may reflect legitimate differences in the studied groups. Comparisons across programs may be somewhat restricted due to differences in such factors as group definitions, testing populations, and test content.

Adverse impact does not reflect test bias if validity research shows that the test is equally valid for relevant groups. “To the extent that members of one group do more poorly on a subset of items that are a legitimate part of the content domain, we would be reluctant to call the discrepancy evidence of bias” (Shepard, 1987). Historically, a regression-based approach has been advocated to evaluate the existence of test bias (e.g., differential prediction). Lack of bias is indicated when the regression line relating the test score [X] and a criterion [Y] is the same for each group. 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. Acquisition of training outcome data from the Services has made it possible to examine the AFQT for differential prediction (test bias). Results of the largest military-sample differential prediction study conducted to date will be provided in the next presentation.

Dr. Manley next addressed special tests that are administered on the ASVAB platform. The Cyber Test is a test of basic computer and information systems knowledge used by all Services. Coding Speed is a speeded test of information processing (assigning code numbers to words) and is only used by the Navy. Dr. Manley indicated that these tests generally exhibited small to moderate effects and were usually as low as or lower than most ASVAB tests. Coding Speed usually had very small effects near 0, but the test may suffer from other issues such as lag time in internet delivery given that it is a speeded test, delivery device and context differences that may affect responses, and the fact that it could suffer from coachability and susceptibility to strategies that result in high scores. He concluded by noting that the potential for AI is not the only consideration when making changes to the ASVAB.

At the end of the briefing, a committee member asked Dr. Manley what is done if there are statistically significant differences in slope and/or intercept, which is likely given the large

sample sizes. Dr. Manley said the next briefing on differential prediction will address those situations. He explained that the sample sizes in that study, however, are smaller because the analyses are conducted by job specialties.

Another committee member thanked Dr. Manley for the work and asked if he would provide his perspective on why subgroup differences sometimes appear to be more or less test-specific (e.g., ASVAB vs. NAEP). Dr. Manley cited slide 31, which showed larger effect sizes on Reading/Verbal on PC, WK, and VE than the NAEP and SAT. He suggested that the Asian population that is interested in military service may not speak English as well as the larger US Asian population; perhaps, they are not a native-US sample. He qualified this, however, as just a guess. He then mentioned that the economy impacts the applicant population: When the economy is bad, more high-aptitude individuals apply to the military. He said, in the end, it is important to keep an eye on the stability of trends over time.

A committee member commented that s/he is glad AI does not reflect test score bias, although the term implies it. S/he then asked, given the differences found, "What should you do? I'm not sure you can do more than what you are already doing." Dr. Velgach said that was an excellent point and that communication is the key. She emphasized the importance and challenge of conveying the fact that AI versus differential prediction means something different to those not in the field.

Dr. Manley said that if the goal is to increase diversity, one would try to decrease the effect size differences, but if the goal is increasing individual performance, then decreasing the effect size differences would not necessarily provide a means to that end. He said, however, that they try to achieve both ends: strike a balance between inclusion and performance.

A committee member asked if DTAC had sufficient first-language data to examine the impact of having English as a second language. Dr. Velgach said that, unfortunately, they do not have those data. She did say, however, that they are going to do an in-depth review of work done in the area and attempt to start collecting the information in a way that it will be useful. She explained that applicants may be wary of providing information, such as first language, which they think will negatively influence their likelihood of qualifying. Following a brief exchange on the difficulty of using proxies (e.g., citizenship) for language, a committee member pointed out the importance of not focusing solely on the mean score level but looking also at the item level. S/he asked when the differential item functioning (DIF) analyses are performed, and Dr. Manley said they are conducted during form development. The committee member asked if there is ongoing DIF work and said s/he was thinking of the impact at the lower end of the distribution; s/he said there may be something going on there. Dr. Velgach asked Dr. Trippe to comment on the methodology, and he said they use the integrated Bayesian model (and cited Hanson) to compare a select number of groups. He said they would make more comparisons if they had sufficient data.

A committee member said that running parallel with Title VII, the IIIA and IIIB categories seem to present a selection and a classification issue, depending on the cut score. S/he said a person can get in with a score of 31, but there are incentives for scoring higher, and though that is not equal to AI in a legal sense, it may be appropriate to aggregate if the only consideration is

selection. Dr. Manley said they are not aggregating at low test scores, though he needed to confirm it. He added that a potential legal issue lies in the opportunity for obtaining better technical jobs; he said there may be a difference in the rate of Black applicants who score above the cut score compared to White applicants for more technical jobs. The committee member suggested a more refined analysis by MOS/rating.

The committee member then asked for clarification on the impact ratio, and Dr. Manley said that was something he would need to explore. He said expanding the criterion space, among other actions, could increase diversity. One could look at other markers of good performance. Another committee member asked about the degree to which the present levels of AI combined with different levels of propensity to join affect the military's ability to develop a force that reflects the diversity levels in the general population. S/he asked, "Given the different levels of propensity, are we getting enough diversity?" Dr. Velgach clarified that the overall enlisted accession population is quite representative compared to the national population overall. She said they have the most difficult time with female representation, which has been a complicated issue to overcome. She solicited recommendations on how to entice that segment of the population. A committee member asked if the problem was the comparison group or propensity to join; that is, should the comparison be to the overall population or to the segment of the population that is propensed to serve? Dr. Velgach said the overall objective is to be representative of the nation. Complementary marketing and advertising programs are being designed to affect propensity. She said they are trying to determine where the biggest effect can be made, whether that is through testing/qualification or communication. Dr. Helland said, because all roles are open to women now, there is a great emphasis on trying to achieve representation.

Another committee member noted that the current discussion was only dealing with the enlisted side and that, at least in the AF, there is not enough representation in the officer ranks. Dr. Velgach said that was an excellent point and that there are special efforts to focus on the officer population, both to increase the pool of the interested applicants and to modernize opportunities for advancement, all the way to the Flag/General Officer level. Dr. Manley pointed out that the requirement for higher test scores has an impact across the range of representation issues in the officer ranks.

13. <u>AFQT Differential Prediction</u> (Tab P)

Dr. Dan Putka, HumRRO, presented the briefing.

> Dr. Putka began by stating that professional standards for psychological measurement recommend test scores used for selection decisions, such as the AFQT, be evaluated for differential prediction, which is defined by the Society for Industrial and Organizational Psychology as "the systematic under- or over-prediction of criterion performance for people belonging to subgroups differentiated by characteristics not relevant to criterion performance." When evaluating differential prediction in the context of personnel selection, these subgroups are often defined by test-takers' race/ethnicity or biological sex. Historically, the literature on differential prediction has been primarily concerned with the under-prediction of performance for the focal subgroup (e.g., Black applicants, female applicants) with respect to positively valenced outcomes such as job performance and training success. If a common regression line results in under-prediction of performance relative to a focal subgroup's regression line, use of the common regression line disadvantages that subgroup. However, if it over-predicts performance relative to the focal subgroup, use of the common regression line does not disadvantage that subgroup. These interpretations are flipped for negatively valenced outcomes such as attrition. If a common regression line results in under-prediction of

attrition relative to a focal subgroup's regression line, use of the common regression line does not disadvantage that subgroup. But if it over-predicts attrition relative to a focal subgroup's regression line, use of the common regression line disadvantages that subgroup.

Dr. Putka continued by explaining that the current study used five years of data on enlisted applicants who completed the ASVAB as part of the ETP. The goal was to evaluate whether AFQT scores exhibited differential prediction for multiple training performance and retention criterion, within multiple Services and across hundreds of enlisted military training jobs/courses. Much past research in civilian and military settings has examined cognitive ability test scores for differential prediction with respect to race/ethnicity, with most focused on White-Black comparisons. General findings have been that use of a common regression line for cognitive ability tests will tend to over-predict performance for Black individuals and under-predict performance for White individuals relative to use of subgroup-specific regression lines. Research results regarding gender have been less consistent, indicating that cognitive ability tests sometimes under-predict females' performance, particularly when college grades are the criterion of interest. But in the Wise et al. (1992) study, where ASVAB technical subtests and technical performance were of primary interest, female performance was either over-predicted or predicted at a comparable level to males.

Despite the amount of research to date, much of the civilian employment research in this area is based on older data (1980s and earlier) and predates the recognition of the methodological concerns raised over the past decade pertaining to the traditional implementation of the Cleary approach. In light of critiques of past differential prediction research, for this study an expanded version of the Cleary approach was adopted that accounts for the presence of range restriction stemming from the use of ASVAB for general enlistment and occupation qualification decisions.

Dr. Putka continued by providing details about the sample and data analyzed in the current study. Data included AFQT scores and demographic information for 1,603,749 individuals who completed the ASVAB as part of the ETP between 1 October 2013 and 30 September 2018. Accession and separation data current through November 2021 were also obtained for the subset of these applicants who accessed into Regular Air Force, Army, Coast Guard, Marine Corps, and Navy, allowing for the calculation of a 36-month attrition criterion. Additional criterion data were also provided for subsets of Air Force, Army, Marine Corps, and Navy accessions from archival data sources from each of those Services.

- Air Force: Awarding course grades for 60K+ Airmen as they completed awarding courses in over 60+ Air Force Specialties (AFS)
- Army: Army-wide job knowledge test (JKT) scores for 38K+ Soldiers from 30+ Military MOS, and MOS-specific JKT scores from 27K+ Soldiers from 10 MOS as they exited initial military training (IMT)
- Marine Corps: Initial military training course graduation indicators (i.e., graduated course without a setback) for 158K+ Marines across 120+ training courses
- Navy: Initial technical training graduation indicators (i.e., graduated without a setback) for 68K+ Sailors across 40+ Ratings

To evaluate whether AFQT scores exhibited differential prediction, and if so, to determine what type of differences they exhibited (i.e., intercept or slope differences), analyses were conducted using an updated version of the Cleary-based approach (described earlier) that accounts for selection-related artifacts raised as concerns in past research. Analyses were conducted for four subgroup contrasts: (1) White non-Hispanic vs. Black non-Hispanic, (2) White non-Hispanic vs. Hispanic White, (3) White non-Hispanic vs. Asian non-Hispanic, (4) Male vs. Female. The magnitude of differences in prediction was estimated for subgroup-specific AFQT regression models by calculating the dMod statistics, and post hoc analyses were performed to estimate the power needed to detect statistically significant differences between the models being compared under the updated Cleary-based approach

Dr. Putka continued by providing an overview of the results of the study.

- Across Services, 664 military occupation/training course-by-subgroup contrast combinations were made to identify evidence of differential prediction for the AFQT for predicting Service-specific criteria. Most combinations examined (79.1%) yielded no statistically significant evidence of differential prediction.
- Consistent with past research, use of a common regression line for the AFQT tended to over-predict performance for Black individuals and under-predict performance for White individuals relative to use of subgroup-specific regression lines. Findings for other focal subgroups reflected a mix of over- and under-prediction when differences were found.
- Across Services, of the 139 military occupation/training course-by-subgroup contrast combinations that exhibited statistically significant evidence of differential prediction for the AFQT, most exhibited small or very small prediction differences at the AFQT Category IIIB lower bound (31) and AFQT Category IIIA lower bound (50).
- Although the AFQT was not designed to predict attrition, relative to the Service-specific criteria, there was more evidence of differential prediction of the AFQT for predicting probabilities of 36-month attrition.
- Prediction differences for attrition were particularly strong for Hispanic White and Asian non-Hispanic Service members relative to White non-Hispanic Service members, with evidence that a common regression line would result in relatively large over-prediction of attrition for Hispanic White and Asian non-Hispanic Service members (particularly in the Army).
- These differences suggest another reason that it is important to consider non-cognitive measures (e.g., TAPAS) if one's objective is to predict first term attrition with a selection/classification measure.

Dr. Putka then showed a series of slides detailing results for each of the Service's performance criteria (Training and Job Knowledge Criteria) as well as 36-month attrition. He then presented suggestions for future research, including gathering stronger criteria data and considering differential prediction for Service-specific ASVAB composites used for occupational qualification. He concluded by prompting the committee for their input on the modified Cleary approach, factors that should be examined to explain the larger instances of over/under prediction, and methods for dealing with limited statistical power observed for occupations with smaller populations.

As Dr. Putka briefed slide 5, he responded to a committee member's earlier comment on sample sizes, explaining that although overall study samples sizes for each service were large, those sample sizes shrink quickly when performing analyses at the job level, and when examining samples sizes for subgroups within jobs. Dr. Putka later noted this has implications for the statistical power of differential prediction analyses conducted within jobs as they were done in this study.

When Dr. Putka was describing future research (slide 34), a committee member asked about the availability of performance measures that are more direct than job knowledge tests. Dr. Putka said that more were available in the source data they had access to, but the JKT and attrition criteria were the ones deemed most suitable for moving forward with for this study. He said he would be glad to address the matter of what is available and why some measures were not used after the meeting.

At the end of the briefing, Dr. Putka mentioned that his team had provided a technical appendix on which he would be glad to receive feedback.

Comments began with a committee member complimenting the effort for being very thoughtful and thorough. S/he then suggested that clustering by related occupations might strengthen the analysis. The committee member also raised the possibility of finding or creating less biased criterion measures. Dr. Putka said that was a good point, but he was not aware of any research that had explicitly examined bias in the criteria used in the study. He did mention they reported observed subgroup differences on the criteria, but such observed differences do not necessarily indicate bias. Regardless, he said they should try to account for that in the future. Dr. Pommerich asked Dr. Kirkendall if the recent criterion measure study had examined bias in criterion measures, but Dr. Kirkendall said it only focused on identifying what measures currently exist. The committee member clarified that no criteria are perfect (e.g., even observed behaviors are contaminated) but recommended remaining sensitive to the issue. Another committee member said JKTs are quite limited in representing the performance domain. Dr. Putka added that supervisor ratings are unreliable (with reliability coefficients typically around .30), though they may be better conceptual representations of performance. The committee member then suggested the possibility of increasing reliability by combining multiple measures. Dr. Putka recognized that might improve reliability (from an item-specific error perspective, but not a rater-specific error perspective), but said it could also muddy the water conceptually. He mentioned fitness and counterproductive work behaviors as other criteria that could be added to the mix. He then raised the issue of the importance of weighting when forming a composite criterion. If a composite is used as a criterion, it would be critical that the weighting of the elements of that composite reflect the policies of the Services in terms of defining overall performance.

A committee member said the study was easily one of the most comprehensive of this nature, part of which is attributable to the sheer amount of data. The committee member then suggested that using a multi-level modeling approach with individuals nested within jobs might result in less bias. Dr. Putka said his team had considered this approach, which he said would be very elegant and parsimonious, but that if his team had done so it would have not been able to account for the presence of selection artifacts (an unresolved issue in the literature). Dr. Velgach mentioned the possibility of looking at class composition and waivers, which would provide more range above and below the cut line. Dr. Putka replied that it would be possible to consider other factors used in selection besides the ASVAB (e.g., various types of waivers), but it would have necessitated a far more complex model than what is typically used in studies of differential prediction.

Another committee member, citing slide 54 (differential prediction outcome summary), asked if the effects were significant at any point along the distribution. Dr. Putka explained that they had constructed visualizations, plotting d_{Mod} distributions to examine effect size along the distribution, but that he did not know of a method to determine significance at specific points along the distribution cleanly.

The discussion closed with a committee member's suggestions on (a) the relevance of native language and (b) the potential examination of results for technical versus non-technical jobs. Dr. Putka said that they had access to indicators of whether English was a second language, but he did not examine such data as part of this effort. Dr. Velgach clarified that their information on whether English was a second language was not reliable.

14. Complex Reasoning (Tab Q)

Mr. Mike Ingerick, HumRRO, presented the briefing.

Mr. Ingerick began by explaining that complex reasoning is a form of non-verbal reasoning that assesses the ability to analyze visual information and solve problems using visual reasoning. It is a meaure of fluid intelligence that has been found to be a strong predictor of training and job success. The 2006 ASVAB Review Panel suggested that DoD consider adding a test of fluid intelligence to better balance the ASVAB's composition (between fluid and crystalized intelligence). The potential benefits of adding a test of fluid intelligence to the ASVAB include improved prediction of training and job success, lower susceptibility to compromise, and increased qualification rates for non-native and non-heritage English speakers. A previous attempt to create a test of fluid intelligence took the form of the Abstract Reasoning Test (ART). It was developed by Susan Embretson of Georgia Tech and had a format similar to Raven's Progressive Matrices items (multiple choice, 6 or 8 response options per item). DTAC commissioned the development of one form (30 items). It was administered (for research purposes) to language training applicants in 2017. The items were found to be relatively easy and time-consuming. Mr. Ingerick then presented two sample ART items.

The objective of the current effort is to develop a complex reasoning testing system to generate items for potential inclusion on the ASVAB. The system should employ a non-proprietary automated item generation (AIG) capability to improve item development efficiency and reduce or eliminate field-testing requirements. Items should be similar to Raven's Progressive Matrices items at a difficulty level appropriate for qualifying military applicants into jobs of varying complexity. The steps in phase I of the effort including evaluating an existing, non-proprietary item generator based on a tool developed by Sandia National Laboratories, developing and refining specifications for the items, developing a pool of items, and pilot testing them with a non-military sample. This will result in refined specifications and a demonstrated proof of concept. Phase II will involve refining/improving the item generation tool, developing additional items using the refined specification and improved tool, and pilot testing those items. These can then be used for research and field testing.

The pilot study was designed to answer several questions.

- Does performance differ by item type (transformation vs. logic)?
- Does performance align with item difficulty specifications?
- Does the number of response options impact performance?
- Does including a "None of these are Correct" option impact performance?
- Does performance differ by gender, race, or ethnicity?
- How much time is needed to complete the items?
- Does completion time differ by item type or response option set?
- Does completion time differ by gender, race, or ethnicity?

Mr. Ingerick continued by presenting a transformation item in which a three-by-three grid is shown with eight of the cells filled with patterns. The examinee is instructed to select (from four response options) the pattern that best completes the grid. Item features include the type, orientation, size, number, and line weighting of the shapes and the direction of the transformations (i.e., vertical, horizontal, diagonal). He then presented a logic item which requires examinees to identify a logic rule that governs the combination of various shapes and select a shape that completes a grid. The pilot design included 24 transformation items ranging from low to high difficulty with the four conditions defined by the number of response options. A "None of these are Correct" alternative was included. A second group received 12 transformation items and 12 logic items also of varying levels of difficulty. There were three conditions defined by the order of item presenation (transformation followed by logic, logic followed by transformation, scrambled).

The sampling plan called for 3,500 civilians aged 18-35 who were U.S. citizens. The measures were (a) one complex reasoning form of 24 items; (b) a post-test questionnaire obtaining demographic information,

perceived difficulty of the items, and test-taking experience; and (c) two attention check items. The measures were administered on the Qualtrics platform with no fixed time limit and a requirement that participants use a desktop or laptop computer. Data were collected from mid-July to early August 2022, with a final sample size of 3,491. Mr. Ingerick then presented a table showing sample demographics by condition. This suggested a fairly even distribution across conditions for each demographic variable (i.e., age, sex, race/ethnicity, education). A second table presented the same information for that portion of the sample that completed the measures in 30 minutes or less and who had a high school degree, a General Educational Development (GED), or less than one year of college.

Mr. Ingerick then showed a series of tables that provided data on unidimensionality, reliability, observed difficulty, group score differences, completion times, and perceived difficulty by condition. The only conditions demonstrating unidimensionality were transformation items with four or eight response options. Of these, the four-response option condition showed somewhat greater difficutly, fewer group differences, and shorter completion times. Based on these findings the recommended item specifications are to include transformation items only with four response options not including a "none of these are correct" option. The item difficulty model and item selection algorithm will be refined to ensure appropriate difficulty levels and minimize group score differences by race-ethnicity where feasible.

Mr. Ingerick concluded by summarizing the proposal for the Phase II pilot. The objective will be to collect data on the refined pool of Complex Reasoning items from a sample representative of military applicants. The results will be used to refine the item specifications and selected a potential pool of complex reasoning items for follow-on research and field testing. The design will include multiple forms of 25 items each, with a subset of common items across forms. A post-test questionnaire similar to that used in the first pilot will be administered along with two attention check items. Participants will be civilians aged 18-35 who are U.S. citizens and have a high school degree, GED, or less than one year of college. Other conditions (method of administration, time limits) will be the same as the first pilot.

At the end of the briefing, a committee member said s/he appreciated the study, to include its potential to reduce AI on non-English speakers, as well as its use of Susan Embretson's work. The committee member said to consider including a fair representation of non-English speakers in the pilot study to increase the potential to validate the test for that population. Mr. Ingerick said that was a great suggestion and that his team will look at who is being invited to address the matter. Another committee member asked whether the team had been inspired more by Susan Embretson's test or drawn more from the Raven's Progressive Matrices. Mr. Ingerick said they had followed the work done by Sandia National Laboratories. Dr. Pommerich added that the pilot sample for phase 2 of the study would be more typical of military applicants.

15. <u>Computational Thinking</u> (Tab R)

Dr. Kimberly Adams, HumRRO, presented the briefing.

Dr. Adams began the presentation by presenting a quote from the National Defense Authorization Act of 2021: "*Not later than one year after the date of the enactment of this Act, the Secretary of Defense shall establish a special purpose test adjunct to the Armed Services Vocational Aptitude Battery test to address computational thinking skills relevant to military applications, including problem decomposition, abstraction, pattern recognition, analytical ability, the identification of variables involved in data representation, and the ability to create algorithms and solution expressions*." Dr. Adams noted that the date for meeting the requirement has been adjusted to October 1, 2024.

Dr. Adams continued by presenting the computational thinking content domains and descriptions.

- Problem decomposition—Breaking down a problem/task into smaller/easier components (e.g., describe a system as a sequence of processes).

- Abstraction—Focusing on most relevant information and ignoring extraneous information to interpret meaning and reduce complexity of a problem/task.
- Pattern recognition—Identifying and using repeated information or patterns to predict outcomes or determine actions for a problem/task.
- Identification of variables involved in data representation—Recognizing how parts of a solution may be reapplied to or eliminated from similar or unique problems/tasks.
- Analytical ability—Inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information for a problem/task.
- Ability to create algorithms and solution expressions—Recognizing and evaluating options against outcomes to simplify or automate processes for efficiency and resource utilization improvements.

She then noted that a measure of Computational Thinking does not currently exist within ASVAB/military testing and the timeframe allowed for implementing one does not support trying to develop one. However, existing ASVAB and other military tests potentially measure content domains underlying the construct. Given that a means to assess the six domains must be operational by October 1, 2024, current plans call for conducting an alignment study to establish a Computational Thinking composite from existing military tests. This will involve:

- Collecting SME judgments linking existing military tests to Computational Thinking content domains.
- Developing estimated correlations among tests either empirically generated when data are available (e.g., among ASVAB subtests) or SME-estimated when data are not available because pairs of tests have never been administered together.
- Using SME judgments and test correlations to derive a Computational Thinking composite score.

Steps involved in this process include (a) updating a literature review to inform Computational Thinking content domain definitions, (b) identifying and recruiting SMES (e.g., I/O psychologists, military testing experts), (c) identifying the military tests of interest, (d) obtaining available correlations among ASVAB and other military tests of interest, (e) developing a linkage exercise and data collection tools, (f) collecting linkage data from SMEs, (g) analyzing and summarizing the results, and (h) providing composite score specifications to DTAC by September 30, 2023.

Dr. Adams continued by listing the tests under consideration, including all the ASVAB subtests, the Complex Reasoning test currently under development, the Cyber Test, Mental Counters, Coding Speed, and Electronics Data Processing. She then presented a potential format for collecting linkage judgments in which the Computational Thinking domains are presented as rows and the various tests under consideration as columns, allowing SMEs to provide input on the relationships between the two. SMEs will be trained in live, virtual sessions with data collected asynchronously. SMEs will be asked to estimate the correlations among each of the six Computational Thinking content domains, between each military test of interest and each content domain, and between pairs of military tests if empirical estimates do not exist. All correlation estimates will be made assuming (a) no range restriction (i.e., applicant population) and (b) the measures of both predictors and Computational Thinking domains are highly reliable.

Materials needed for the alignment study include definitions of each of the Computational Thinking domains, SME-provided estimates of the correlations between the domains, test blueprints for each of the military tests with 2-3 non-operational sample items from each domain, and estimated correlations between the military tests (empirical or SME provided). Dr. Adams continued by presenting a timeline for alignment study activities.

The objectives of the evaluation study will be to validate the Computational Thinking composite score derived from the synthetic equation developed in Phase I using a Computational Thinking score based on administration of the marker instruments using military applicants and recruits or a similar population. The composite score will be evaluated in regard to score distributions, subgroup differences, and other pertinent outcomes to be determined. Steps will include (a) specifying the data collection method and participant recruitment strategy, (b) modifying the platform or other tools required to administer the tests, (c) recruiting

participants, (d) administering the tests, (e) computing correlations between the Computational Thinking composite score and scores from the marker instruments, and (f) conducting analyses to evaluate other pertinent outcomes. Dr. Adams then presented a tentative timeline for conducting this work. Another chart was shown comparing two options for data collection using either applicants/recruits or a random sample of individuals representative of ASVAB examinees along with the procedures to be followed, the strengths of each, and factors that must be taken into consideration for each approach. She concluded by outlining the next steps in the process, including finalizing alignment study plans, completing Institutional Review Board and/or Exemption Determination Official paperwork, preparing data collection tools, and recruiting SMEs for the alignment study. She concluded by soliciting committee input on both the alignment and evaluation studies.

A committee member asked Dr. Adams if she would be conducting the alignment study at the item level or the subtest level. Dr. Adams said that it would be done at the subtest level because that is where scores are computed. She said doing the study – and creating the measure – at the item level would require a major change to the testing program.

Another committee member asked why one of the existing marker instruments could not be used as the new measure. Dr. Adams said no existing marker instrument represents all six content domains and most are for lower grades, such that their content is insufficiently advanced. Because of this, she said, they may have to administer more than one marker test in the pilot study.

In regard to data collection strategies (slide 20), Dr. Adams said she would consider the committee's earlier comments on increasing the representation of non-English speakers in the pilot study sample.

At the end of the briefing, a committee member suggested looking into the work of Zach Hambrick, who s/he said has developed a similar measure, though it is not exactly a computational reasoning test.

16. <u>ASVAB CEP Update</u> (Tab S)

Ms. Kate McLean and Dr. Jan Bayer, Written, LLC., presented the briefing.

Ms. McLean began the presentation by showing various tables documenting aspects of the program over the past several years.

- In 2022, 607,324 students in 12,907 schools took part in the program, generating 494,981 leads to the Military Services, all of which exceed the numbers during the previous two years that were impacted by the pandemic. There was an error on slide 4 for 2021 participating students (corrected number is 311,465).
- In 2022, approximately 82% of students who participated in the program took the P&P version of the ASVAB, which demonstrates a small but consistent decrease from previous years (2021 85%, 2020 89%, 2019 91%).
- In 2022, approximately 36% of the leads generated were in the Category I to IIIA aptitude range.
- In 2022, 14,234 students used their CEP ASVAB scores to enlist in the military, which was the smallest number in the past five years.
- CEP website utilization figures indicate there were more users in 2021-2022 than in previous years (1,037,704), with significant increases in tablet/mobile visitors (591,726).
- Careers in the Military (CTM) utilization figures showed increases in 2022 over previous years in number of users, returning users, page views, and number of tablet/mobile users.

Ms. McLean then provided figures regarding inquiries received through both the CEP and CTM websites. She continued by indicating that more than 35 states have enacted legislation requiring college and career readiness indicators. Monitoring state activities regarding the CEP shows that various options have been promoted including allowing CEP scores to serve as an indicator of military or career readiness for federal reporting purposes, requiring or endorsing the CEP for career exploration purposes, or simply mentioning the ASVAB CEP as an option for career exploration. Ms. McLean then presented a map of the United States annotated to show which states have taken such steps.

Ms. McLean continued by presenting a list of CEP marketing activities, including both print and digital media, and well as a list of national events at which the CEP has been promoted through attendance, exhibits, and/or presentations. She went on to explain that the National Board of Certified Counselors (NBCC) and Affiliates is a not-for-profit certification organization that grants Continuing Education (CE) hours to interested counselors. In October 2020, the ASVAB CEP became a CE provider to offer single programs for NBCC continuing education hours on ASVAB CEP components. Monthly sessions are offered, and attendees can earn one hour of free NBCC credit.

Dr. Bayer then turned to ConCEP, which is a new platform developed for MEPCOM to consolidate access to all ASVAB CEP applications and modernize processes and systems surrounding scheduling, marketing, and communications. She detailed various features of the application, including the fact that it operates on a web-based nationwide secure network and provides role-based views with business rules and workflows. A version of ConCEP that addresses scheduling was launched in the summer of 2022 and ongoing improvements have been implemented. Plans are to migrate the system to the government cloud at MEPCOM, prepare to apply for an Authority to Operate, and perform software changes as needed.

Dr. Bayer then presented a brief history of occupational crosswalks. The Military and Occupational Database (MOTD), established in 1984 by the DMDC, included a crosswalk between military occupational codes and various federal taxonomies. A Military Careers Publication was originally produced by DMDC but was shifted to the Joint Advertising Market Research & Studies group while retaining input from both CEP and DMDC personnel. Various other military-civilian career crosswalks have been developed over time, including the Department of Labor's Veterans Opportunity to Work (VOW) study and the Service Credentialing Opportunities On-Line (COOL). The ASVAB CEP crosswalk expansion began in 2006 using a combination of automated and manual analyses. These efforts expanded the number of military-civilian linkages from 881 to over 21,000.

Recent efforts include updating the military-civilian crosswalk to incorporate the new O*NET-Standard Occupational Classification 2019 taxonomy. There is now at least one military match for each O*NET data-level occupation to provide CEP users with related opportunities in the military. Within the O*NET program, these updates were made available via O*NET OnLine, My Next Move for Veterans, O*NET web services, and downloadable crosswalk files. Seven new high-level military careers are now available on the CTM website, including Data Analysts, Business Operations Specialists, and Search and Rescue Specialists. The Crosswalk of Occupational Resources application is for internal use only and allows for various functions such as analyzing skill importance ratings related to aptitude, assigning interest codes, and analyzing career paths. Recent efforts have also included establishing linkages between O*NET codes and US Office of Personnel Management federal occupations, which will be available on the CEP website in July 2023.

Dr. Bayer then turned the discussion to the upcoming revision of the ASVAB CEP website, knows as CEP 2.0. She noted that legislative requirements included updates for schools on ways to satisfy the Every Student Succeeds Act mandates, including establishing career plans and encouraging mentor involvement. She then presented a series of CEP 2.0 screenshots showing how students can create a plan, invite a mentor to view the plan, and allow the mentor to provide feedback.

Dr. Bayer ended the presentation by discussing PTI training, the goal of which is to standardize the process through which CEP participants are provided information about the meaning of their scores and ways they can be used with other program resources to explore future careers and pursuits. A train-the-trainer

methodology is employed, including three steps: (1) take part in virtual training; (2) attend in-person training; and (3) nominate, observe, and evaluate others. In addition to trainee feedback, metrics to gauge success include increased utilization of program websites, increased testing numbers, and training attendance. With the introduction of CEP 2.0, recertification training will be offered in April and May of 2023 at three locations, including an enhanced train-the-trainer component. In addition, Test Coordinator training will be made available virtually based on a training needs assessment that included interviews with SMEs, job description evaluations, and lists of competencies and content areas.

At the start of the briefing, Dr. Velgach emphasized the importance of the ASVAB CEP – the program as a whole – to military recruiting. She explained that it is one of the main ways the military is able to access high school students. She said the program offers something for everyone and enables effective engagement about the military as well as the full spectrum of occupational opportunities. She said she wanted the committee to provide input on the best way to structure the program to maintain and increase access to students.

As Ms. McLean briefed slide 5, Dr. Velgach explained that, unlike the ETP, most of the CEP tests are P&P because of bandwidth limitations in schools.

On CTM utilization (slide 9), a committee member asked if users see all the branches in which a job is available, as well as branch differentiating information to assist with fit. Ms. McLean said users can see which Services offer each job, but they can also drill down by Service. She said, within each Service, users can search by career level. Dr. Velgach added that the site includes military and civilian jobs. The committee member asked if DTAC knows how users decide on a Service. Dr. Helland said that they collect data on why recruits choose their Service in a recruit survey.

When Ms. McLean explained that the number of score requests had doubled from approximately 5,000 to 10,000, a committee member asked if requests to bring the CEP to a school had increased as well. Ms. McLean said, remarkably, it had decreased, most likely due to the addition of a question that asked "when" they would like to have the test administered. Dr. Velgach asked the team to think about whether the "when" question should be retained. Dr. Pommerich asked if COVID-19 was still a factor. Ms. McLean said it was, in that part of the increase from 2021 to 2022 was likely due to the decreasing concern about COVID. Mr. Davis said it took some time for MEPCOM to get back into schools, explaining that they lost access in March 2020 and schools still apply extra scrutiny in determining who can visit.

On legislative emphasis (slide 12), a committee member commented on the fact that Colorado allows the ASVAB to be used to meet graduation requirements. Ms. McLean said Indiana does the same thing, and that this has increased test use by over 400%. She said increasing state engagement while maintaining the capability to handle demand is a challenge. Dr. Velgach said additional funding is not allocated solely because a state is using the test more frequently, which causes manpower issues.

A brief discussion on marketing highlighted the need to retain the name "ASVAB" in order to maintain the level of credibility associated with the test.

Dr. Bayer identified two issues related to PTI proficiency training (slide 37): students thought their percentile scores were their percent correct scores and believed they had to join the military if the participated in a PTI.

At the end of the briefing, a committee member asked – from a parent and counselor perspective – who holds the testing coordinators accountable. Mr. Davis said the ESSs are responsible for the program, the test administrators (TAs) administer the test, and the training function belongs to the test control officers. He said the test control officers are also responsible for security. He said that an ESS or TA may provide a PTI, and Dr. Velgach added to that list recruiters who have been certified. Dr. Bayer said the situation is challenging because the responsibility is so distributed: the ESSs work in the local MEPS and the recruiters work for the Services. She said the ESSs coordinate PTIs, and everything is supposed to go through them so that they can ensure consistency. She said the test control officer, for whom the TAs work, is also responsible.

The committee member responded by saying the CEP appeared to be a great program, but from a cultural perspective, some may not want their kids to take the test because of a fear they might have to go to war. S/he said promoting the test in a well-balanced manner is very important because the program is so much more than just a military tool. Dr. Bayer said there has been a lot of marketing of the test as just that: more than military. She said, in the PTIs, students are told that military service is not required. Further, recruiters are told that it is okay to recruit, but not during a PTI. She said the whole endeavor is a challenge and wondered if the counselors were disseminating information effectively. A committee member said that some kids take the test instead of the ACT or SAT because it is free. Dr. Velgach said they promote the test in that way (i.e., that it is a viable means of meeting State and school requirements), but that there are still many teachers and school counselors that are hesitant to let the military into schools. Dr. Helland raised the point that using the term "free" in association with the program seems to devalue it.

A committee member asked if the program has competition; that is, is there anything comparable or better? Dr. Bayer said some districts use programs such as Naviance® or other fee-based programs. She said she emphasizes that the CEP is unique because it exposes users to *all* the options, to include Federal civilian options, which are currently being added. Dr. Velgach explained that the CEP is the only program that includes aptitude *and* interest components. She said the amount of work done by this very small team is almost inconceivable and that they have done a superb job.

A committee member asked how long students can access the site after taking the test. Dr. Bayer said 2 years, or through June 30th at least two years after testing. The committee member asked if the CEP team was on social media and if they used paid ads. The answer was yes. The committee member asked if the test could be administered on a mobile device application, but the response was negative. CDR Pyne explained how students are able to use social media to share information about the test, including their careers of interest. Dr. Bayer said there are strategically placed opportunities on the website for sharing information. She said each month they use a toolkit to promote test participation and that she would appreciate the committee's ideas on how to improve that function. A committee responded with three recommendations: (a) use social media to facilitate a culture of interest in schools, because peers matter; (b) emphasize the focus on exploring jobs and work, as opposed to college (which is an emphasis of

Naviance®); and (c) stress the "whole-person" nature of the assessment, how it measures the breadth of who you are, identifying cognitive *and* non-cognitive strengths and weaknesses. The committee member said a little of these types of information goes a long way.

Dr. Velgach said it might be helpful to arrange for the committee members to work through the website; she mentioned the professional development piece is incorporated into the website. Dr. Bayer said she could arrange a virtual meeting to make that happen.

Dr. Velgach then asked the committee for their thoughts on the ramifications of states that are wanting to make the ASVAB mandatory, such as Texas. She mentioned, for example, that this would drive participation by students who did not want to participate, which would lead to scores that are not representative. Dr. Bayer said it would be challenging. Mr. Davis said MEPCOM did not have the manpower to cover something like that. Dr. Velgach led the discussion back to the matter of implications for test validity and asked the committee to start thinking about it. Dr. Pommerich said they would need to consider how States are using the scores; she said that DTAC only validates the test for certain uses. A committee member asked where one would take the test if his/her school did not offer it. Dr. Bayer said anyone can contact their local MEPS to find another school or even a homeschool group that may offer the test. Mr. Davis said school counselors should also have that information.

A committee member commented on the need to persuade students not only to take the test, but to take it seriously. S/he said explaining that it can identify pathways – realistic pathways – forward to good careers (e.g., through exposing students to role model examples) should be an emphasis, which is something typical career counseling lacks. Mr. Davis mentioned that, when school populations do not want to participate, MEPCOM asks the counselors to conduct activities to encourage participation.

The discussion closed with a committee member asking about bringing the program to community and technical colleges, and Mr. Davis said MEPCOM does offer a program for that purpose.

17. <u>Future Topics</u> (Tab T)

Dr. Mary Pommerich, DTAC, presented the briefing.

Dr. Pommerich presented a list of potential topics for future DACMPT meetings, as follows:

- ASVAB development
  - Item development procedures
  - Item writing guidelines and tools
- CAT-ASVAB/Form development methodology
  - CAT-ASVAB modernization evaluation
  - Using machine learning methods to streamline form assembly process
  - Item calibration sample size reduction study
- Unproctored testing
  - P*i*CAT/VTest updates
  - APT
- Military compatibility assessment
- Adding new non-cognitive measures

    - TAPAS/personality measures
        - Joint-Service TAPAS effort
        - Service TAPAS efforts (Army, Air Force/Space Force, Marine Corps, Navy)
        - TAPAS validity framework
- Interest measures
- CEP updates
    - Website demos
- Adding new cognitive tests/composites
    - Cyber Test
    - Mental Counters
    - Complex Reasoning update
    - Computational Thinking update
- Web and cloud efforts
    - Device expansion
    - *i*CAT expansion
- Next Generation Testing
    - ASVAB evaluations
    - High school curriculum study
- Impact of COVID/score trends for ETP & CEP
- ASVAB validity
    - ASVAB/AFQT validity framework updates
    - Criterion domain/performance metrics
- Social media effort

As Dr. Pommerich talked through the future topics ideas, Dr. Velgach said the next meeting should certainly cover the item development methodology and equating. Committee members suggested addressing non-cognitive measures, the high school curriculum study, and the TAPAS validity framework. Additionally, topics on Non-Native English Speakers and the CEP Website were recommended. One committee member recommended that topics should be selected with an eye toward the big picture of integrating measures into the master plan for selection and classification.

18. Public Comments

At the end of the second day, Dr. Velgach opened the floor to public comments and asked participants to limit their comments to 5 minutes per person.

Commander Mike Pyne (Navy) expressed interest in how the TAPAS might be used to accept recruits under the condition that their occupation would be decided at a later date. Dr. Velgach mentioned the Army's Future Soldiers program as a means of helping recruits, especially with academic knowledge and preparation to meet the Army's fitness standards. She said the other Services are looking at implementing similar programs.

Mr. Scheidt mentioned the potential of tests, such as the TAPAS, to expand the number of jobs for which people qualify. He explained that those with AFQT scores of around 31 might not technically qualify for certain classifications, though they are able to perform those jobs capably. Ms. Katie Gunther (Space Force) referred to an effort to bring professionals into the officer ranks to address shortages in various areas.

There were no additional comments from the public.

In closing the meeting, Dr. Velgach thanked the presenters, saying that it is rare to see such organized and detailed presentations. She also thanked the audience for their participation, saying that it was great to hear how the systems and products work in the operational environment. A committee member thanked Dr. Velgach for the privilege of being on the DACMPT. The committee member said the testing program is remarkable and no one else has a research program that is even close to being able to provide the support that this program allows. Dr. Velgach thanked the committee members for their time and willingness to give back to the military. She stressed the value of receiving outside feedback to evaluating plans and outcomes in an objective manner.

Tab A

LIST OF ATTENDEES

Defense Advisory Committee on Military Personnel Testing (DACMPT)
December 15-16, 2022

Name	Position	Organization
Dr. Nancy Tippins	Owner and Manager	DACMPT, Nancy Tippins Group, LLC
Dr. Sonia Esquivel	Professor	DACMPT, US Air Force Academy
Dr. Won-Chan Lee	Professor	DACMPT, University of Iowa
Dr. Osvaldo Morera	Professor	DACMPT, University of Texas El Paso
Dr. Fred Oswald	Professor	DACMPT, Rice University
Dr. April Zenisky	Associate Professor	DACMPT, University of Massachusetts, Amherst
Dr. Sofiya Velgach	Designated Federal Officer (attendance req'd by FACA)	Office of Accession Policy (AP)
Dr. Katherine Helland	Director-inbound	AP
Mr. Christopher Graves	Senior Staff Scientist	Human Resources Research Organization (HumRRO)
Ms. Sachi Phillips	Project Manager	HumRRO
Dr. Mary Pommerich	Director	Defense Testing and Assessment Center (DTAC)
Dr. Tia Fechter	Personnel Research Psychologist	DTAC
Dr. Greg Manley	Personnel Research Psychologist	DTAC
Dr. Ping Yin	Personnel Research Psychologist	DTAC
Mr. Matthew Scheidt	Supervisory Survey Statistician	US Marine Corps, MANDR Affairs
Ms. Katie Gunther	Chief, Studies and Analysis, Guardian Talent Management	US Space Force
SGM Alan Myers	Senior Retention & Accessions Policy Manager	US Army HQDA, G1
Dr. Tonia Heffner	Chief, Selection and Assignment Research Unit	US Army Research Institute (ARI)

Dr. Cristina Kirkendall	Research Psychologist	ARI
Dr. Amanda Mouton	Personnel Research Psychologist	US Air Force
Dr. Sophie Romay	Senior Personnel Research Psychologist	Air Force Personnel Center
Mr. Ken Schwartz	Chief, Testing and Survey Policy	Air Force Personnel Policy
Mr. Robert Tiegs	Testing Director	US Military Entrance Processing Command (USMEPCOM)
Mr. Wayne Crook	Command TCO/Management Analyst	USMEPCOM
Mr. David Davis	Chief, Testing Division	USMEPCOM
Mr. Jaime Clayton	Enlistment Testing Program Manager	USMEPCOM
CDR Mike Pyne	Chief of Naval Personnel	US Navy, OPNAV N132
Dr. Matthew Trippe	Program Manager	HumRRO
Dr. Scott Oppler	Principal Scientist	HumRRO
Dr. Claire Vincent	Program Manager	HumRRO
Dr. Deirdre Knapp	Principal Scientist	HumRRO
Dr. Michelle Hardoin	Program Manager	HumRRO
Dr. Kimberly Adams	Program Manager	HumRRO
Mr. Mike Ingerick	Business Development Director	HumRRO
Ms. Pamela Baumer	Research Scientist	HumRRO
Dr. Furong Gao	Senior Staff Scientist	HumRRO
Dr. Dan Putka	Principal Scientist	HumRRO
Dr. Tom Waterbury	Research Scientist	HumRRO
Dr. Elizabeth Waterbury	Senior Scientist	HumRRO
Ms. Kate McLean	Owner	Written, LLC
Dr. Jan Bayer	Vice President, Research and Analysis	Written, LLC

Tab B

AGENDA

Defense Advisory Committee on Military Personnel Testing (DACMPT)
December 15-16, 2022

December 15, 2022 (Pacific Time)

8:30 a.m. – 8:45 a.m.	Welcome and Opening Remarks	Dr. Sofiya Velgach, OASD(M&RA)/AP
8:45 a.m. – 9:15 a.m.	Accession Policy Introduction	Dr. Sofiya Velgach OASD(M&RA)/AP
9:15 a.m. – 10:00 a.m.	New Member Briefing	Dr. Mary Pommerich OPA/DTAC
10:00 a.m. – 10:15 a.m.	*Break*	
10:15 a.m. – 11:00 a.m.	Milestones Briefing	Dr. Mary Pommerich OPA/DTAC*
11:00 a.m. – 12:00 p.m.	ASVAB Form development	Dr. Matt Trippe HumRRO
12:00 p.m. – 1:30 p.m.	*Lunch*	
1:30 p.m. – 3:30 p.m.	Next Generation ASVAB/Testing	Dr. Mary Pommerich (OPA/DTAC)
	a. ASVAB Evaluation Plan	Dr. Mary Pommerich (OPA/DTAC)
	b. ASVAB /AFQT Validity Framework	Dr. Deirdre Knapp (HumRRO)
	c. Training Relevance Survey	Dr. Scott Oppler (HumRRO)
	d. Focus Groups	Dr. Kimberly Adams (HumRRO)
3:30 p.m. – 3:45 p.m.	*Break*	
3:45 p.m. – 4:15 p.m.	Norming Requirements/Plans	Dr. Pamela Baumer HumRRO
4:15 p.m. – 5:00 p.m.	Device Expansion Plans	Dr. Tia Fechter, OPA/DTAC
5:00 p.m. – 5:15 p.m.	*Public Comments*	

December 16, 2022 (Pacific Time)

8:30 a.m. – 9:30 a.m.	ASVAB Adverse Impact	Dr. Greg Manley OPA/DTAC
9:30 a.m. – 10:30 a.m.	AFQT Differential Prediction	Dr. Dan Putka HumRRO
10:30 a.m. – 10:45 a.m.	*Break*	
10:45 a.m. – 11:15 a.m.	Complex Reasoning	Dr. Mike Ingerick HumRRO
11:15 a.m. – 11:45 a.m.	Computational Thinking	Dr. Kimberly Adams HumRRO
11:45 a.m. – 12:00 p.m.	*Break*	
12:00 p.m. – 12:45 p.m.	ASVAB CEP Update	Ms. Kate McLean Dr. Jan Bayer Written, LLC.
12:45 p.m. – 1:00 p.m.	Future Topics –	Mary Pommerich OPA/DTAC
1:00 p.m. – 1:15 p.m.	*Public Comments*	
1:15 p.m. – 1:30 p.m.	Closing Comments	Dr. Nancy Tippins Chair

ABBREVIATIONS KEY:
ASVAB - Armed Services Vocational Aptitude Battery
ASVAB CEP - ASVAB Career Exploration Program, student testing program provided free to high schools nation-wide to help students develop career exploration skills and used by recruiters to identify potential applicants for enlistment
AFQT - Armed Forces Qualification Test
CAT - Computer Adaptive Testing
HumRRO - Human Resources Research Organization
OASD(M&RA)/AP - Office of the Assistant Secretary of Defense (Manpower & Reserve Affairs)/Accession Policy
OPA/DTAC - Office of People Analytics/Defense Testing and Assessment Center

Tab C

LIST OF ACRONYMS

3PL	Three-Parameter Logistic Model
ACT	American College Testing Test
AFCT	Armed Forces Classification Test
AFQT	Armed Forces Qualification Test
AFS	Air Force Specialties
AI	Automotive Information
AI	Adverse Impact
AIG	Automatic Item Generation
AO	Assembling Objects
AP	Accession Policy
APT	AFQT Prediction Test
AR	Arithmetic Reasoning
ART	Abstract Reasoning Test
AS	Auto & Shop
ASVAB	Armed Services Vocational Aptitude Battery
BLS	Bureau of Labor Statistics
CAT-ASVAB	Computerized Adaptive Testing ASVAB
CDF	Cumulative Distribution Functions
CDR	Commander
CE	Continuing Education
CEP	Career Exploration Program
CFR	Code of Federal Regulations
COOL	Credentialing Opportunities On-Line
COVID-19	2019 Coronavirus
CTM	Careers in the Military
DACMPT	Defense Advisory Committee on Military Personnel Testing
DASD-MPP	Deputy Assistant Secretary of Defense for Military Personnel Policy
DFO	Designated Federal Officer
DIF	Differential Item Functioning
DMDC	Defense Manpower Data Center
DoD	Department of Defense
DoL	Department of Labor
DTAC	Defense Testing and Assessment Center
EI	Electronics Information
ESS	Education Services Specialists
ETP	Enlistment Testing Program
FACA	Federal Advisory Committee Act
FLRI	Foreign Language Recruiting Initiative
FY	Fiscal Year
FYI	Find Your Interests

GED	General Educational Development
GS	General Science
HumRRO	Human Resources Research Organization
IMT	Initial Military Training
IRT	Item Response Theory
*i*CAT	Internet version of the CAT-ASVAB
IMT	Initial Military Training
JAMRS	Joint Advertising Market Research and Studies
JKT	Job Knowledge Test
LD	Local Dependence
MAPWG	Military Accession Policy Working Group
MC	Mechanical Comprehension
MEPCOM	Military Entrance Processing Command
MEPS	Military Entrance Processing Stations
METs	Military Enlistment Testing Site
MK	Mathematics Knowledge
MLSP	Modified Stocking-Lord Procedure
MOS	Military Occupational Specialty
MOTD	Military and Occupational Database
MSLP	Modified Stocking-Lord Procedure
M&RA	Manpower & Reserve Affairs
NAEP	National Assessment of Educational Progress
NBCC	National Board of Certified Counselors
NDAA	National Defense Authorization Act
NLSY	National Longitudinal Survey of Youth
OASD	Office of the Assistant Secretary of Defense
OPA	Office of People Analytics
OSD P&R	Under Secretary of Defense for Personnel and Readiness
P&P	Paper-and-Pencil
PC	Paragraph Comprehension
P*i*CAT	Pending Internet Computerized Adaptive Test
PTI	Post-Test Interpretation
QR	Quantitative Reasoning
SAT	Scholastic Aptitude Test
SGM	Sergeant Major
SES	Socioeconomic Status
SI	Shop Information
SME	Subject Matter Expert
TAPAS	Tailored Adaptive Personality Assessment System
TIF	Test Information Functions
TOA	Theory of Action
USC	U. S. Code
VA	Verbal Ability

VE	Verbal Composite
VOW	Veterans Opportunity to Work
VTest	Verification Test
WK	Word Knowledge

Tab D

NANCY T TIPPINS
GROUP

March 11, 2023

Katherine Helland, Ph.D.
Director, Accession Policy
Accession Policy
Room 3D1066
4000 Defense Pentagon
Washington DC 20301-40003

Dear Dr. Helland,

The Defense Advisory Committee on Personnel Testing (DACMPT) is pleased to provide this report of our meeting of December 15-16, 2022, in Monterey, California. Below, we provide summaries and recommendations from the DACMPT emanating from that meeting, which was highly informative, productive, and well-facilitated. The setting promoted interactions among the presenters and Committee members as well as other participants in the meeting who are interested in military personnel testing. In addition to myself, DACMPT Committee members are Drs. April Zenisky, Fred Oswald, Won-Chan Lee, Osvaldo Morera, and Sonia Esquivel.

The meeting began with opening remarks from Dr. Sofiya Velgach (Assistant Director for Testing Standards, Office of the Under Secretary of Defense for Personnel and Readiness/M&RA/MPP(AP))

and Dr. Nancy Tippins (Chair of the DACMPT). In addition to the DACMPT Committee members, representatives from HumRRO, staff from the Defense Testing and Assessment Center (DTAC), and various military units were present. Dr. Velgach introduced you (Dr. Katherine Helland) because you had been recently appointed to the position of Director of Access Policy and were in attendance. All but one member of the DACMPT participated in person. Several of the presenters on the meeting agenda made their remarks virtually. For the most part, the internet connection worked reliably, although occasionally, one of the virtual participants could not hear the proceedings well .

The DACMPT report and recommendations follow, in the order of the meeting agenda.

Accession Policy Introduction

Dr. Sofiya Velgach provided an introduction to Accession Policy (AP). She began her comments with a statement of the Accession Policy mission: “Develop, review, and analyze policies, resources, and plans for Services’ enlisted recruiting and officer commissioning programs.” After presenting the organizational structure, Dr. Velgach highlighted the activities of the Accession Policy portfolio in seven categories: USMEPCOM, Recruiting, Testing, Officer Programs, Personnel Security, Medical, and Other. She then reviewed the recruiting challenges faced by the services with the Committee.

Recommendations

Because of so many new members to the DACMPT, the Committee found the overview particularly helpful and would like to be regularly updated on Accession Policy’s activities, including the challenges it faces in accomplishing its mission.

New Member Briefing

Dr. Mary Pommerich provided the next presentation. The presentation objective was to orient the new members of the DACMPT by presenting information regarding the military testing program, starting with the organizations involved: Defense Testing and Assessment Center (DTAC), United States Military Entrance Processing Command (USMEPCOM), Accession Policy (AP), and the Defense Advisory Committee on Military Personnel Testing (DACMPT). After explaining the role of each entity and their organizational relationships, she provided detailed information on the ASVAB, Special Purpose Tests, and Language Tests and on ongoing research efforts.

Recommendations

The DACMPT appreciated the detailed information and wishes to be updated on changes to the testing programs as well as the results of the research efforts being conducted.

ASVAB R&D Efforts: Milestones and Project Schedules

Dr. Mary Pommerich provided an overview of the major R&D initiatives underway and updated the schedule for all the projects. This work includes developing new Computerized Adaptive Testing (CAT)-ASVAB item pools, creating new pencil and paper ASVAB forms, evaluating the methodologies for the CAT-ASVAB and paper-form development, as well as looking ahead to activities associated with next-generation ASVAB research (including adverse impact, differential prediction, training relevance, high school curricula, focus group studies, validity frameworks, and norming investigations). Other R&D activities focus on evaluating new cognitive tests and composite scores for the ASVAB (complex reasoning, computational thinking, cyber test, and mental counters), adding noncognitive measures to selection and classification as well as for compatibility assessment, reviewing and maintaining all materials for the career exploration program (CEP), and continuing work to develop test availability through web/cloud delivery and device expansion.

Recommendations

The DACMPT appreciated the scope of research on the ASVAB and other cognitive and non-cognitive measures and the efforts to improve the career exploration program and delivery of tests. The Committee supported the ongoing review of current high school course content, curriculum, standards, and instructional methods to ensure that the next-generation ASVAB is aligned to high school content, particularly with respect to courses likely to be taken by individuals inclined to join the Services. The Committee requests a curated list of technical reports (and access to them as appropriate) and updates regarding progress on this research.

CAT-ASVAB Pool and P&P Form Development

Dr. Matthew Trippe, Research Manager and Project Director for the ASVAB Forms Project at HumRRO, provided an overview of the CAT-ASVAB item pools and the P&P-ASVAB forms. Dr. Trippe described the approach used for the development of item pools in the CAT mode and the key processes involved in the production of paper forms. Much work has been done on local item dependence and item enemy identification and mitigation, and considerable efforts have been undertaken to ensure that the psychometric processes for pool assembly, given exposure control constraints, and for equating are adhering to professional standards in these areas.

Dr. Trippe also described technical challenges associated with the Auto & Shop (AS) test, in which the AS test is scaled as a composite for CAT-ASVAB and separately for P&P-ASVAB (for historical/operational reasons). Analyses conducted to investigate strategies for equating these items resulted in the use of a modified Stocking-Lord procedure (MSLP).

Another technical challenge that has been researched involved the difference between the structure of items in the Paragraph Comprehension (PC) test between CAT and P&P, where the CAT-ASVAB has 1 item per stimulus (10 items total) and the P&P-ASVAB has 5 questions per stimulus (3 stimuli). Testing time must be comparable between modes, meaning the P&P-ASVAB cannot consist of 15 items (and consequently 15 stimuli). Research suggests that a 10-item P&P-ASVAB PC test may be optimal given measurement and operational constraints, and a recommendation will be finalized in the coming months.

Recommendations

The DACMPT inquired about the transformation steps taken in terms of equating to understand better the processes used and to ensure that variability was not being introduced as a consequence of methodology. More information regarding these steps and the results is requested. Additional information on the efforts to detect and manage multidimensionality in data from CAT-ASVAB forms is also requested. The DACMPT also requests more information about the nature of the PC Test stimuli (length, content focus on informational vs. literary reading), given the research to meet operational constraints and ensure comparability between P&P and CAT.

Next Generation ASVAB/Testing – Evaluation Plan

Dr. Mary Pommerich provided background and updated Committee on the history, status, and plans for the next generation of ASVAB and special tests administered on the ASVAB platform in the military's Enlistment Testing Program (ETP). In 2005–2006, the ASVAB underwent a systematic review, resulting in recommendations for improvements and enhancements to the military's ETP. Some research and development efforts are still underway. The contents of the ASVAB have not yet changed, due in part to the lack of consensus on the philosophy of the ASVAB, logistical difficulties, and concerns about insufficient resources. For next generation testing effort, DTAC hopes to work through some of these issues. Relying on an ASVAB Stakeholder Advisory Committee and focus group information, DTAC plans to develop, refine, and evaluate the tests in the ASVAB and special tests. The goal of this work is to improve selection into the military and classification into military occupational specialties while conforming to testing standards. Total testing time will be a key consideration if additional tests are incorporated into the ASVAB due to administration constraints. Dr. Pommerich then presented a progress report on Next Generation Testing, including Mental Counters, Cyber Test, Complex Reasoning, and TAPAS.

Recommendations

The DACMPT asked how DTAC defined improvements in selection (e.g., increases in validity or satisfaction). The answer will require another look at the philosophy or purpose of the ASVAB. The DACMPT recommends careful consideration of the criteria for "improvement."

Committee members recognized the diversity of needs among stakeholders. For example, military trainers are generally pleased with the current tests because new recruits succeed during training. At the same time, recruiters want a test that will qualify more people and allow them to meet their recruiting missions. Although a completely shared vision for the ASVAB is likely impossible, there are no major complaints, DTAC is hoping to meet most of the stakeholders' goals. The DACMPT encourages

continued efforts to evaluate stakeholder perceptions and to educate them on the compromises that must be made.

The primary concern about testing time comes from MEPCOM, which prefers to complete all testing in a single day to avoid overnight stays, and not from applicants. Committee members discussed the issue of the length of the tests and briefly explored alternatives such as changing the CAT stop rules, moving item seeding requirements from proctored testing to VTest administrations, employing psychometric refinements, using a multidimensional approach (e.g., multidimensional IRT), and initiating a taxonomy content review to identify redundancies. Although the tests are already short, the DACMPT recommends that DTAC continue to explore various ways to shorten the length of time required for administering the ASVAB and special tests.

The Committee also discussed applicant perceptions of the ASVAB. The available data were collected from individuals who had taken the ASVAB but had not yet completed training and did not include high school students taking the CEP or applicants who were not accepted. The DACMPT encourages efforts to understand s broader range of applicant reactions to the ASVAB.

ASVAB /AFQT Validity Framework

Dr. Deirdre Knapp explained the Validity Framework for the ASVAB and the AFQT. The ASVAB/AFQT uses Theory of Action (TOA) as an organizing framework for validity, which reflects an interpretative argument (inferences that scores are intended to support) and a validity argument (evidence that provides justification for supporting the inferences in the interpretative argument). This framework was part of a challenging intensive process that involves developing solid interpretive arguments (i.e., ones that are clear, cohesive, comprehensive).

TOA is being applied to the ASVAB/AFQT for enlisted selection, for TAPAS enlisted selection and classification, and for an in-progress project on ASVAB classification. The TOA for the AFQT was centered around general cognitive ability (a composite of AFQT scores) in its prediction of performance (both training and job performance) for the purpose of making effective selection decisions. This claim is generally supported by the literature, but then operationalized by AFQT and performance data in this project, and ultimately supported in their results.

Recommendations

The Committee agreed that TOA was applied very successfully in the AFQT selection context presented in developing, justifying, and empirically supporting the claims that were tested. Committee members appreciated how TOA-based validation efforts can usefully evolve over time. No validity evidence is static, and the TOA approach allows the body of validation work to be revised as the literature changes, and in light of different stakeholder purposes. ASVAB for classification may be more useful when average scores are higher because scores are less correlated (Legree, et al., 1996[1]). The DACMPT recommends continued use of the TOA as an organizing framework for validity.

[1] Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence, 23*(1), 45–57. https://doi.org/10.1016/S0160-2896(96)80005-5

Training Relevance Survey

Dr Scott Oppler of HumRRO presented the results of the ASVAB Training Relevance Survey, which was co-authored by Dr. Kimberly Adams. The purpose of the survey was to evaluate the relevance of ASVAB content as well as other test content/constructs to success in entry-level military training. The approach involved administering a survey to technical training course developers and instructors to collect ratings of relevance to success in technical training for entry-level military occupations and knowledge levels.

The results of the survey support the relevance of the majority of content categories included on the ASVAB to training performance. The primary exception may be General Science (GS) for which only a small number of content areas, primarily in the Physical Sciences, received ratings in support of their relevance to training performance. Several tests (i.e., Assembling Objects, Cyber Test, the six Computational Thinking areas) may be underutilized, given the relatively high relevancy ratings for various training courses.

Dr. Oppler solicited feedback from the DACMPT on how DTAC can use these results as they consider the Next Generation ASVAB as well as any other comments they may have. Committee members raised questions regarding the handling of multiple ratings of courses and tests by the same person, prior exposure to the constructs measured by the Cyber Test, and restriction of range in these results.

Recommendations

The DACMPT has no recommendations for improvement with the administration and analyses of the relevance survey.

Next-Generation Testing Stakeholder Focus Group Study

Dr. Kimberly Adams of HumRRO reported on the methodologies for and findings of stakeholder group focus groups to explore perspectives about the ASVAB and ETP (Enlistment Testing Program) and generally to inform conversations around the next-generation ASVAB and ETP. Three realms of stakeholders were identified (military/DoD, Education, and Examinee), and within each group, recruitment of study participants occurred within multiple roles. About 300 individuals participated across 42 realm-specific focus groups (251 individuals in military/DoD, 31 in Education, and 20 Examinees). The interview protocol inquired about ASVAB/ETP experience in terms of likes and dislikes and solicited ideas for suggested changes as well as potential barriers to change. DoD/military participants indicated that overall, participants perceive that the ASVAB generally works as a predictor of entry-level MOS success; it improves person-job fit decision-making; and it measures a wide range of skills. Less-positive impressions of ASVAB/ETP that were identified included a lack of measures of one's potential and/or ability to learn, the perception of outdated or not applicable content, and limited value for prediction across all aspects of performance.

Recommendations

The DACMPT asked about the representation of the study participants relative to the populations. Demographic information about the participants was limited. Consequently, the sample did not meet strict sampling conditions. The DACMPT recommends that future focus group ensure adequate representation of all critical groups.

Norming Requirements/Plans

Ms. Pamela Baumer began the briefing by summarizing past efforts to norm the ASVAB and recent evaluations of the norms by comparing changes in performance on various national tests such as the National Assessment of Education Progress (NAEP) and Scholastic Aptitude Test (SAT) to identify shifts in performance. Negligible differences in scores in Categories IIIB, IVC, and V were found between 1980 and 1997, but no differences were found between the 1997 percentages and later ones. Although there does not appear to be an urgent need for re-norming, a working panel will be convened to explore approaches to re-norming and set the criteria for when norming efforts should occur. Discussions with the DoL regarding coordination with efforts to re-norm the 2026 National Longitudinal Survey of Youth (NLSY) have begun. The focus will be on the composition of the norming group rather than the age of the normative information.

Ms. Baumer prompted the Committee for their thoughts on several issues, including estimating ability changes for groups when there are no available data, defining meaningful change in AFQT scores that should trigger a new norming effort, identifying other methods for collecting the needed norming data, and considering the impact of COVID-19 on educational achievement.

Committee members discussed the value of using standard scores instead of percentile scores, the impact of labor market conditions on scores, the appropriate frequency of evaluating changes in scores, the imputation of score changes from past years' data, and the effects of COVID on test scores.

Recommendations

One Committee member asked for a plot of trend results for AFQT scores.

The Committee discussed the possible effects of COVID on test scores, noting that some groups were more affected than others. The DACMPT recommends that efforts to re-norm should be deferred until the effects of COVID on propensity to serve have abated.

The DACMPT recommended that DTAC be sensitive to changes resulting from more vulnerable groups being differentially affected and wait until more time has elapsed before initiating a major re-norming effort. In addition, the methodology used for re-norming the ACT and SAT should be considered as plans to re-norm the ASVAB are developed.

The Committee also explored the development of norms based on the applicant pool instead of the customary approach of using the entire population. The DACMPT recommends that the DTAC consider the relative advantages and disadvantages of each approach before deciding which approach to use.

Device Expansion Plans

Dr. Tia Fechter of DTAC, Office of People Analytics (OPA), presented an update on an initiative to explore the feasibility of expanding the number and variety of devices that can be used to complete unproctored assessments, including a next generation user interface (UI) that is responsive to device/operating system/browser variations. Prior research suggests that subtest scores for ASVAB may be comparable across devices if the device used is familiar and the UI is responsive to device types. Dr. Fechter also reviewed operational implementation considerations, focusing on *who* should take ASVAB on *which* device for *what* purpose(s).

Recommendations

The DACMPT asked about research on the interaction of item features and device variability to determine if different performance was observed for different items and tests when delivered on different devices, taking into account interactions among familiarity with the device, the task to be performed, response action, and device. Another question was raised about mode comparability research and the studies that were done or planned to ensure comparability of results across devices, operating systems, and browsers.

The DACMPT made several recommendations regarding future research into alternate devices and their effects on test scores. Continuing research in this area should focus on differential analyses, as well as interaction effects that may impact dropping items from tests and/or evolving technologies (hardware and software). Data on the nature of the task, information on how the content is displayed, and the test taker's knowledge of moving around the screen should be collected and incorporated into the research.

ASVAB Adverse Impact

Dr. Greg Manley of OPA/Defense Testing and Assessment Center provided a summary of the adverse impact analyses for the ASVAB. In this research, adverse impact was assessed by comparing the ratio of minority and majority pass rates and using the four-fifths rule to determine impact (i.e., ratios below .80 indicate adverse impact). ASVAB scores of 31 and 50 were used to define qualified individuals. (A score of 31 or above is a common score for allowing the examinee to enlist in the services; a score of 50 or above allows for the use of enlistment incentives.) In addition, standardized mean differences were the effect sizes presented. The data presented were collected from October 1, 2020, through September 30, 2021. Due to COVID-19, there were issues with missing demographic data, but reported effect size differences were comparable to previous assessments of adverse impact. In addition, the adverse impact of the ASVAB and that of two other large-scale testing programs involving high school seniors (SAT 2016 scores for college bound high school seniors in Math and Reading and the NAEP 2019 scores for Reading, Math, and Science) were compared. Overall, the magnitude of adverse impact observed was consistent over time; however, there was some variability in adverse impact based on the test being evaluated and the group providing the data.

Recommendation

The DACMPT recommends regular analyses of adverse impact and exploration of potential reasons for differences in test performance to aid in promoting diverse accessions into the military services. Future assessments of adverse impact should also consider whether English is the examinee's first language.

AFQT Differential Prediction Summary

Dr. Dan Putka of HumRRO gave a presentation on evaluating AFQT scores for differential prediction across groups. The use of a single regression line would disadvantage focal groups (e.g., women, African Americans, Hispanics and Asian Americans) if it resulted in under-prediction of outcomes such as training success or job performance. Data from 1,603,749 individuals who completed the ASVAB as part of the ETP between October 1, 2013 and September 30, 2018, and for whom job knowledge test scores and 36-month attrition information were available were analyzed.

To evaluate differential prediction, Dr. Putka and his team modified Cleary's three-step approach. First, the outcome variable was regressed on the AFQT predictor score. Second, group membership was entered as a second predictor to determine if a single regression line was acceptable versus group-specific parallel regression lines (intercept invariance). Third, the interaction between group and AFQT scores was entered to see whether parallel lines were acceptable or whether slopes needed to differ to

account for predictive AFQT-outcome relationships for each group (slope invariance). Effect sizes based on the d_{mod} statistic were used to quantify the amount of differential prediction, and Cohen's heuristic rules for 'small', 'medium' and 'large' effect sizes were used to classify the amount of differential prediction.

Dr. Putka examined 664 military occupation/training courses and found that across all possible group comparisons, almost 80% of them provided no statistical evidence of differential prediction. Consistent with past literature, the common regression equation tended to under-predict for White examinees, relative to African American examinees. Of the 139 analyses that revealed differential prediction, effect size estimates revealed mostly small or very small differences at the cut-scores of 31 (Category IIIB) and 50 (Category IIIA). Dr. Putka and his team also demonstrated that although the ASVAB was not designed to predict 36-month attrition, there was evidence of differential prediction for this outcome variable when a common regression equation would have led to over-prediction of attrition for Hispanic and Asian American examinees.

Recommendations

Dr. Putka requested input from the DACMPT in three areas: the modified Cleary approach to assess differential prediction, other factors that may explain overprediction and underprediction, and approaches for dealing with limited power for analyses involving occupations with small sample sizes. Committee members noted that overprediction was expected and asked questions regarding combinations of outcome measures, the effect of the scores of individuals who did not make it into the study, the use of multilevel modeling for these multi-group analyses, other ways to probe differential prediction, (e.g., using the Johnson-Neyman regions of significance approach; Preacher, Curran & Bauer, 2006[2]), and the use of multilevel modeling to address selection artifacts and comparisons involving technical and non-technical occupations.

After discussion of the approach taken and the available data for such analyses, the DACMPT made several suggestions regarding modifications to this research that might be considered: using performance measures that are broader and more direct than job knowledge tests, clustering related jobs or sorting jobs into technical and non-technical positions, using multilevel modeling as an analytic approach be considered going forward, and evaluating the effect of the test taker's native language. Despite these suggestions, the DACMPT is aware that the data needed for these initiatives may not exist at all, may not be reliably collected, or may not be available for a sufficient sample of test takers.

Complex Reasoning

Dr. Mike Ingerick defined complex reasoning (CR) as nonverbal, using visual information, and requiring visual reasoning. CR has been shown to be a strong predictor of training and job success, while maintaining greater test security than other cognitive measures. The original CR measure that was developed by Susan Embretson looks much like Raven's Progressive Matrices (RPM). In early testing, the original CR measure was found to be both too easy and too time consuming. Therefore, the new proposed system generates automated items with properties found to be more psychometrically appropriate and effective for military classification. A Phase 1 pilot design comprised 24 CR items that involve (a) transformation in terms of shape (e.g., type, orientation, size, number, shading) and their

[2] Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437-448.

arrangement (e.g., vertical, horizontal, diagonal) and (b) logic that defines the pattern (e.g., AND, OR, exclusive OR). Items were administered in an untimed format to a large non-military sample of young adults who were race-diverse and gender-balanced. Results revealed scores were most favorable for the items that only required transformation and lacked a "none of these options are correct" (NOTAC) response. These items showed high reliability (α = .75 or higher) and small subgroup race and gender differences overall, in addition to taking less time to complete. Phase 2 will test military applicants, use transformation items with no NOTAC response options, and select items with appropriate difficulty and lower subgroup differences.

Recommendations

The DACMPT valued the development of a complex reasoning measure because such a measure is lacking in the ASVAB and virtually all jobs in the military require complex reasoning. Complex reasoning measures require very little verbal ability and therefore may be fairer to applicants, so long as they are familiar with this type of test. The DACMPT suggested that future research consider including non-English speakers in the pilot study to increase the potential to validate the test for those populations.

Computational Thinking

Dr. Kimberly Adams led the discussion on the computational thinking test. The National Defense Authorization Act (NDAA) for Fiscal Year 2021 (HR 6395), Section 594 seeks to "establish a special-purpose test adjunct to ASVAB to address computational thinking skills relevant to military applications, including problem decomposition, abstraction, pattern recognition, analytical ability, the identification of variables involved in data preparation, and the ability to create algorithms and solution expressions." The deadline based on Section 581 in the FY 2022 NDAA is October 1, 2024, which prevents engaging in a full validation process. In terms of application, the computational thinking measure might better fit into an occupation-specific classification process than the overall selection process. The NDAA leaves this option open; it also leaves open how one might weight the six computational thinking domains (overall and within specific occupations). Computational Thinking development will involve conducting a literature review, conceptually and empirically correlating the computational thinking measure with existing tests (ASVAB, complex reasoning, etc.), and gathering SME judgments about linkages to other measures. Scores will be analyzed in terms of score distributions, subgroup differences, correlations with related measures, and other outcomes. Different data collection strategies were proposed to this end, such that this effort is both practical and coordinated.

Recommendations

The DACMPT supports the development of the computational thinking measure via a composite and the plans for doing so. Many jobs in the military have increased requirements to develop, engage in, and solve technological problems. Consequently, the development and implementation of a computational thinking measure will likely improve military classification. More specifically, the Committee suggested increasing the representation of non-English speakers in the pilot study sample and reviewing the work of Zach Hambrick, who has developed a similar measure.

ASVAB CEP

Ms. Kate McLean and Dr. Jane Bayer provided an overview of the Career Exploration Program (CEP) from 2018 to 2022. Despite minimal funding and the challenges of COVID, the CEP is widely used and produces a number of sound leads for enlistment. The number of iCAT administrations is steadily increasing. Website utilization is high, with over 7M page views for the ASVAB site and over 3M for the

Careers in the Military site in 2021-2022. More than 35 states have included some kind of "college-and-career readiness" indicator, such as using ASVAB scores for a high school graduation credential, for an indicator of graduation fitness, or for an indicator of military and/or career readiness for state/federal reporting purposes. Some states have incorporated the ASVAB into their career exploration processes. The CEP is widely promoted through presentations to career development professionals and continuing education credits for career counselors. The new ConCEP platform, developed for the Military Entrance Processing Command (MEPCOM) and launched in the Summer of 2022 consolidates access to all ASVAB CEP applications and modernizes processes and systems surrounding scheduling, marketing, and communications. Work continues on maintaining and updating military-civilian occupational crosswalks. CEP 2.0 helps meet legislative requirements and satisfy the Every Student Succeeds Act (ESSA).

Recommendations

Much of the discussion and resulting recommendations revolved around ways to encourage participation in the CEP. The DACMPT suggested that the “Bring ASVAB CEP to your school” program be looked at closely to determine if the scheduling forum has pushed people away since 2019 and if the demographic questions on the forum should be revised. The DACMPT also suggested that students be assigned an identification code (e.g., pseudo name or number) to reduce the concerns about military service. Other suggestions included using social media to facilitate a culture of interest in schools, emphasizing the focus on exploring jobs and work as opposed to college and stressing the “whole-person” nature of the assessment. The Committee also felt that strong student testimonies placed on the homepage might engage more users and should be considered. Other efforts to engage more users include working jointly with programs like Upward Bound and offering the program to undeclared freshmen in college and those in the TRIO program. YouTube videos that are aligned with the topics in the “Student Articles” would also be helpful. Understanding other programs in high school that compete with the ASVAB-CEP could help direct marketing efforts, and the use of social media tools such as Kahoot could enable better connections among educators, students, and the military.

No major concerns were uncovered; however, the DACMPT would like to see more information regarding the Army’s success in using CEP scores for enlistment. The DACMPT also requests that the following questions be addressed in future meetings:

- Should ASVAB-CEP be mandatory for high school students, and what will be the ramifications for the military services?
- What methods will best persuade students to take the ASVAB-CEP and take it seriously?
- How can the military promote, “Do you know people like you who took the ASVAB-CEP”?
- How should non-cognitive measures be incorporated into the selection and classification programs?
- How does the content in high school curricula align with the ASVAB, and what are implication for changes to either or both?

Finally, the Committee endorses the idea of the Committee members working through the website to better understand the program.

Future Topics

Dr. Pommerich presented a list of potential future topics and led a discussion of them with the DACMPT.

Recommendations

The DACMPT recommends future meetings incorporate briefings on the following topics:

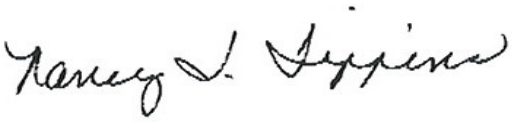
- Item development and equating methodology
- Non-cognitive measures
- The high school curriculum study
- The TAPAS validity framework
- Non-native English Speakers and test performance
- CEP Website
- Integrating measures into the master plan for selection and classification.

Summary

Full funding of the ASVAB and associated programs over the years has promoted significant progress on key goals and activities of Accession Policy. The current ASVAB program is inarguably one of the most sophisticated testing programs in the world. The research that supports its applications represents the state of the art in large, high-stakes selection programs. To continue to allow the DoD to meet its goals for high-quality accession, training, and force readiness, the DACMPT strongly encourages the federal government to maintain full funding for the ASVAB program and associated projects.

Overall, the meeting was informative and useful. The DACMPT appreciates the efforts of Accession Policy and DTAC staff and the research staff of each of the services as well as the consultants who provide their services to the DoD. As always, the DACMPT is interested in supporting these efforts, as they provide strong, well-informed, and timely justification for the intended interpretations and uses of the ASVAB. We look forward to our next meeting.

Sincerely,



Nancy T. Tippins, Ph.D.
Principal, The Nancy T. Tippins Group, LLC
Chair, Defense Advisory Committee on Military Personnel Testing