OFFICE OF PEOPLE ANALYTICS

Device Expansion Plans

Tia Fechter & Mary Pommerich
Defense Testing & Assessment Center

DAC Meeting
December 15, 2022

OVERVIEW

- Goals & Impact
- History of Device Evaluation
- Operational Implementation Considerations
- Recommendations & Progress
- Test Monitoring Plan
- DAC Feedback
- Backup Slides

OPA
OFFICE OF PEOPLE ANALYTICS

GOALS & IMPACT

- 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 (i.e., PiCAT, APT).
- Allow for more flexibility for ASVAB administration to reduce time spent in MEPS and increase number of applicants.
- Develop a Next Generation user interface that incorporates a Responsive Design approach, which automatically formats the test display to alternative devices.

HISTORY OF DEVICE EVALUATION: QUESTIONS & FINDINGS

- **Does device differentially impact examinee performance (score; response time) on ASVAB subtests?**
  - The specific device an examinee uses to take the ASVAB does not significantly impact test scores.
  - In general, examinees use less time responding to items on alternative devices in comparison to the Dell XPS Notebook; however, differences were not practically significant (<30 seconds for a full length ASVAB subtest).*

*Note: Previous studies have shown that examinees are provided sufficient time for responding to test items on currently allowed administration devices such as the XPS.

- **Does device familiarity differentially impact examinee performance on ASVAB subtests?**
  - Examinees perform better on the ASVAB when they are familiar with the device they use.

HISTORY OF DEVICE EVALUATION: QUESTIONS & FINDINGS

- **Does device differentially impact item difficulty?**
  - Most likely not.
  - 7.4% of possible comparisons (1,116) were flagged for DIF (based on device type), where a Type I error rate of 5% suggests that many are likely to be unexplainable.

- **Are there item features (e.g., inclusion of graphic) that interact with the device, which increase the probability that item difficulty is differentially impacted?**
  - None that were explicitly studied
  - Two possible post-hoc reasons for DIF for some items were classified into
    - Preference for an item display for some items with certain features
    - Ease of looking up answers via an internet search

HISTORY OF DEVICE EVALUATION: QUESTIONS & FINDINGS

- **Are there any interactions of device familiarity and historically disadvantaged subgroups that impact examinee score performance on ASVAB subtests?**
  - Yes, but only in one comparison.
  - A positive effect of familiarity with device was heightened for the lower SES subgroup on the Math Knowledge (MK) subtest—that is, the score differences widen for those unfamiliar with the device they used to take the ASVAB.
- **Overall, based on these findings, ASVAB subtest scores among applicants should be comparable regardless of device used to take the tests so long as the examinee uses a device that is familiar to him/her AND the test delivery application is designed to be responsive to a variety of device types.**

OPERATIONAL IMPLEMENTATION CONSIDERATIONS

Who should take the ASVAB on which mobile device and for what purpose?

- **Who?**
  - Applicants testing at home?
  - Applicants testing at MEPS/METS?
  - Students testing in CEP/schools?
- **Which mobile devices?**
  - Test-taker owned & maintained?
  - DoD owned & maintained?
  - School owned & maintained?
- **What purpose?**
  - Unproctored APT—a screener where score is not score of record
  - Unproctored PiCAT— verification required for enlistment
  - Proctored at MEPS/METS—score of record
  - Proctored at high schools—score of record

Some considerations: Compromise from test-taker-owned device via screenshots; maintenance costs of DoD-owned devices; score effects associated with testing on unfamiliar devices

OPA
OFFICE OF PEOPLE ANALYTICS

RECOMMENDATIONS & PROGRESS

- **Design a test delivery application for the ASVAB that is responsive to a variety of device types**
- **Make operational implementation decisions prior to moving forward with device expansion**
  - Focus on devices commonly used by those eligible for military service or by students in high schools
  - Smartphones will be limited to low stakes testing (i.e., APT) and ASVAB tests where scores must be verified (i.e., PiCAT) to mitigate security concerns
  - Expand browser & operating system options, where feasible

RECOMMENDATIONS & PROGRESS

- **Make operational implementation decisions prior to moving forward with device expansion**
  - Three-phase implementation plan (for unproctored tests)

Phase	Device	Operating System	Browser				
			Chrome	Safari	Edge	Firefox	Samsung
1	Desktops & laptops	Windows	X		X	X	
		macOS	X	X	X	X	
		ChromeOS	X				
2	Tablets & touchscreen technology	Windows	X		X	X	
		iPadOS		X			
		Android	X		X	X	X
3	Smart phones	Android					X
		iOS		X			

OPA
OFFICE OF PEOPLE ANALYTICS

RECOMMENDATIONS & PROGRESS

- **Allow examinees to choose a device they are familiar with to take the ASVAB**
  - Note that in the September 2020 meeting of the DACMPT, a committee member cautioned that allowing device choice may not necessarily maximize the benefits of familiarity, citing the tangential research literature on students' lack of ability to make optimal decisions regarding which items to respond to (basing their decisions on content familiarity and not necessarily on the cognitive task they can perform best).
  - Mitigations included providing clear advice to examinees as to which devices and device specifications would be most appropriate (or optimal) for test taking.

RECOMMENDATIONS & PROGRESS

- **Develop a test-monitoring plan that tracks operational performance differences (scores & response time) between device types**
  - Recently completed and incorporates the next two recommendations
  - Included a literature review on unproctored mobile testing, resulting in additional recommendations to DTAC
- **Develop a data-collection tool that reports device features (e.g., screen size, browser type and version, device type, etc.) for post-test monitoring and analysis**
- **Develop and implement a post-test questionnaire intended to measure barriers to optimal performance**

TEST-MONITORING PLAN: DATA CURRENTLY CAPTURED

- SSN (for iCAT, PiCAT)
- Test Form
- Test Date
- Platform (e.g., CEP, PiCAT)
- Overall time to complete
- Completion times for each subtest
- Failure to complete individual subtests
- Item response times for each subtest subtest
- Failure recoveries & failure recovery times
- Subtest scores
- Expired/terminated tests
- IDs of item administered
- Demographics
  - Age
  - Gender
  - Race/Ethnicity

OPA
OFFICE OF PEOPLE ANALYTICS

TEST-MONITORING PLAN: PROPOSED* ADDITIONAL DATA TO CAPTURE

- Operating System
- Browser version
- Responses to optional pre-/post-test questionnaire for unproctored tests
  - Device specifications (e.g., model, age of device)
  - Familiarity with device/operating system/browser used
  - Test-taking environment (including distraction conditions)
  - Motivation level
  - Perceptions of fatigue
  - Connectivity issues

* Note that not all may be feasible or practical to collect/implement.

TEST-MONITORING PLAN: OUTCOME MEASURES TO MONITOR

- Data from tests administered on alternate devices will be compared to those on standard devices (e.g., desktop computers)

Level of Analysis	Outcome Variables	
	Performance	Response Time
Item	Difficulty Indices	
Subtest	Scaled Scores	
Composite	Scaled Scores	
Battery		

OPA
OFFICE OF PEOPLE ANALYTICS

TEST-MONITORING PLAN: VARIABLES OF INTEREST*

- Device/operating system/browser
  - Including examinee familiarity
- Test administration conditions
  - Distractions
  - Technical issues
- Demographics
- Features or characteristics of items or tests
- Others to be determined based on feedback (e.g., education level) from implementation team and consultants

* Note that not all may be feasible or practical to collect/implement.

TEST-MONITORING PLAN: TYPES OF ANALYSES

- Various Generalized Linear Models (GLM)
- Differential Item Functioning (DIF)

TEST-MONITORING PLAN

- Frequency of monitoring will depend on sample size conditions
  - Develop sample size targets to initiate each analysis
    - Statistical power
    - Type I error
    - Expediency required to answer questions
- Ongoing monitoring
  - Until there is sufficient evidence that device type does not significantly impact performance
  - Resumed when device types are changed or expanded
- Isolation of triggers for performance differences (e.g., screen scrolling requirements) & remediation plan

TEST-MONITORING PLAN: ADDITIONAL LITERATURE REVIEW RECOMMENDATIONS

- Seek advisement from consultants with in-depth knowledge of the practice of administering high-stakes assessments on mobile devices to ensure all implications have been considered.
  - We are actively forming an advisory group to discuss operational implementation.
- Systematically monitor data on an ongoing basis & create template for reporting outcomes to be used by decision-makers to promptly address any necessary actions.

DAC FEEDBACK

- Any other considerations we should address with the implementation panel?
- Overall thoughts on way forward?

OPA

OFFICE OF PEOPLE ANALYTICS

BACKUP SLIDES

ACRONYMS & DEVICES USED

- **Testing Programs, Assessments, Composite Scores, & Subtests**
  - AFCT: Armed Forces Classification Test
  - AFQT: Armed Forces Qualification Test (a composite score from ASVAB)
  - APT: AFQT Predictor Test
  - ASVAB: Armed Services Vocational Aptitude Battery
  - CEP: Career Exploration Program
  - ETP: Enlisted Testing Program
  - iCAT: Internet Computer Adaptive Test
  - MK: Mathematical Knowledge (an ASVAB subtest)
  - PiCAT: Pending Internet Computer Adaptive Test

- **Assessment Sites**
  - MEPS: Military Entrance Processing Station
  - METS: Military Entrance Test Site

OPA
OFFICE OF PEOPLE ANALYTICS

ACRONYMS & DEVICES USED

- **Devices**
  - XPS: Dell XPS 13 (the device that served as the control condition)
  - Apple MacBook Pro
  - Apple iPhone XS
  - Apple iPad Pro
  - Samsung Galaxy Tab A
  - Samsung Galaxy S9+
  - Dell Chromebook 3380

- **Other**
  - DIF: Differential Item Functioning
  - DoD: Department of Defense
  - GLM: Generalized Linear Model
  - SES: Socio-Economic Status

PARTICIPANT DEVICE FAMILIARITY

Which electronic devices are you comfortable using? Please select all that apply.

- ○ Dell XPS Laptop
- ○ Apple MacBook Pro Laptop
- ○ Apple iPhone
- ○ Apple iPad Pro
- ○ Samsung Galaxy Tablet
- ○ Samsung Galaxy Smartphone
- ○ Dell Chromebook Laptop

OPA
OFFICE OF PEOPLE ANALYTICS