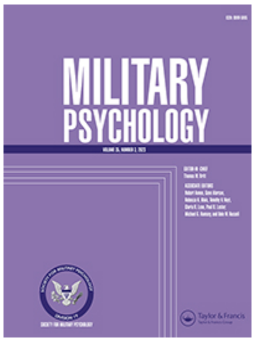

Military Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmlp20

Development of a serious gaming approach for cyber aptitude assessment

Jaclyn Martin-Kowal, Brenton Wiernik, Thomas R. Carretta & Michael D. Coovert

MILITARY PSYCHOLOGY
https://doi.org/10.1080/08995605.2021.1984740

Development of a serious gaming approach for cyber aptitude assessment

Jaclyn Martin-Kowal[a,b], Brenton Wiernik[a,c], Thomas R. Carretta[d], and Michael D. Coovert[a,c]

[a]MDC & Associates, Tampa, Florida, USA; [b]Personnel Decisions Research Institute, Arlington, Virginia, USA; [c]Department of Psychology, University of South Florida, Tampa, Florida, USA; [d]Air Force Research Laboratory, Wright-Patterson AFB, Ohio, USA

ABSTRACT

Numerous traditional assessments have been developed to determine suitability of US military recruits for cyber careers. Cyber career field managers expressed a concern there may be well-qualified candidates that lack cyber knowledge, and therefore are not identified with knowledge-based tests. Technological advances such as serious gaming may provide opportunities to assess constructs traditional methods do not effectively measure. The purpose of this effort was to identify potential gains in validity that could be achieved beyond traditional methods through the use of serious games for several cyber jobs (both for enlisted and officer positions). Throughout this phase of research, an extensive literature review of military and civilian assessments targeted cyber occupations. Then, military subject matter experts in these career fields provided input and guidance (e.g., focus on aptitudes and traits as knowledge and skill are rapidly outdated). A gap analysis between all measures of such constructs identified a short list of candidates for measurement in a serious game. A survey of 800 airmen in the 1N4X1A, 3D1X2 and 17DEX/SX career fields was conducted; 290 respondents identified six constructs to be the focus for serious game assessment. The game was developed, and constructs validated on a sample chosen to model Air Force enlisted recruits. Additional psychometric data from enlistees and cyber trainees will be gathered once COVID-19 restrictions are lifted.

ARTICLE HISTORY
Received 31 March 2021
Accepted 16 September 2021

KEYWORDS
Cyber aptitude; serious games; personnel measurement

What is the public significance of this article?—This study advances research on the use of serious games for psychological measurement with the demonstrated early success of a game developed to assess constructs important to performance in cyber career fields.

Introduction

In 2006, the US Air Force (USAF) announced that cyberspace would become a new mission area. In the same year, a panel of experts in cognitive psychology, test development, and psychometrics suggested the addition of a test of information/communications technology to the Armed Services Vocational Aptitude Battery (ASVAB) (Drasgow et al., 2006). In 2010, the Department of Defense announced the establishment of the US Cyber Command (McMichael, 2010) whose mission was to coordinate offensive and defensive cyber activities. The demand for high quality cyber/information technology personnel is great, as is competition for them among industry, the government, and military (Gould, 2013).

Several assessments have been developed or proposed for classification of military personnel into cyber careers. Most of these involve measurement of cognitive abilities (e.g., ASVAB, Electronic Data Processing Test), technical knowledge (e.g., Cyber Test), or personality traits (e.g., Self-Description Inventory, Tailored Adaptive Personality Assessment System). These assessments have been the focus of extensive research across a broad range of career fields, and employ common testing methods (e.g., paper-and-pencil or computer-administered) and approaches to validation. One exception is the proposed use of serious games, or assessments that incorporate the use of video game elements to improve user experience and user engagement, in cyber assessment.

Although traditional cyber aptitude assessments have demonstrated predictive validity against initial military training performance (M. D. Trippe et al., 2014), cyber career field managers have expressed concerns that their focus on crystalized intelligence and cyber knowledge may fail to identify well-qualified applicants who lack prior cyber experience. Furthermore, in May of 2019, the President of the United States issued an executive

CONTACT Jaclyn Martin-Kowal jaclyn.martin28@gmail.com 8234 Kippis Road, Millersville, MD 21108.
Supplemental data for this article can be accessed on the publisher's website.

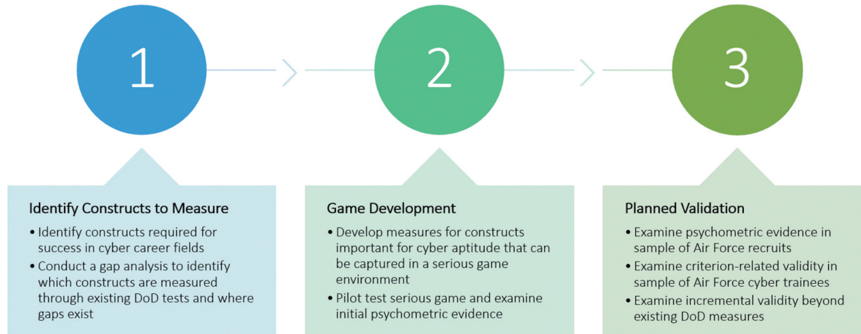

Figure 1. Project phases and objectives.

order directing federal agencies to "identify a list of cybersecurity aptitude assessments for agencies to use in identifying current employees with the potential to acquire cybersecurity skills," highlighting that the issue is of national importance (Exec. Order No. 13870, 2019). With the demand for cyber workers outpacing the supply of cyber workers (Bureau of Labor Statistics, US Department of Labor, 2021), employers need to increase their available applicant pool through pivoting focus from technical knowledge and certifications to more distal predictors of job performance, such as abilities and traits.

The scientific literature on abilities and traits important for cyber-security job performance is primarily conceptual and theoretical in nature with limited predictive validation data currently available. The non-technical abilities and traits proposed to be important for cyber-security job performance are systems thinking ability, analytical thinking, active or continuous learning, communication, integrity/civic duty, time management/selective attention/vigilance, and attention to detail (Dawson & Thomson, 2018; Dreibelbis et al., 2018; Jose et al., 2016).

Some of these constructs are difficult to measure with traditional self-report methods. Aside from allowing for the measurement of more complex constructs, serious games increase psychological fidelity, which allows for greater face validity of the assessment and can increase candidates' perceptions of fairness in the selection process. Serious games could also increase engagement for applicants, which could result in less applicant attrition and more positive candidate reactions when compared to traditional assessment. Moreover, serious games could allow for the measurement of multiple constructs simultaneously, which may ultimately reduce testing time.

Proponents for the use of serious games in selection and classification argue that technology-enhanced assessments increase face validity and lead to positive examinee reactions, perceptions of realism, and fairness. This may be especially true for occupations with a heavy technological component (Tippins, 2011), such as military cyber careers. A gaming approach to personnel assessment is more entertaining and engaging than traditional assessment methods, making potentially boring tasks more interesting (Handler, 2013), and may better match the testing preferences of newer generations of military applicants who have had access to complex, realistic, and sophisticated games for most or all of their lives.

This paper summarizes ongoing research on the development of a serious gaming approach for constructs related to cyber aptitude. The first section covers the first phase of research, which involved conducting an extensive review of archival Air Force materials and scientific literature, and interviews with subject matter experts (SMEs). This review yielded 62 important constructs related to cyber aptitude. A gap analysis identifying existing measures available for the important constructs, and input through a survey of 800 airmen narrowed the list down to six critical constructs. Next, the game development and psychometric evidence collected during the second phase of research is summarized. Finally, a summary of the work and limitations is provided. See Figure 1 for an overview of the phases of research and associated objectives.

Identification of constructs critical for cyber aptitude

Identification of the aptitudes and traits[1] required for competency in various Air Force cyber career fields (enlisted: 3D0X2 – Cyber Systems Operations; 3D0X3 – Cyber Surety; 1B4X1 – Cyber Warfare Operations; 1N4X1A – Digital Network Analyst; officer: 17DX – Network Operations; 17SX – Cyber Space Operations) began with a review of the military research on testing and assessment and SME input. Next, scientific outlets were considered to garner a complete picture of the required individual attributes. A gap analysis was done to focus future work on measures of those attributes that require assessment development. As a validation check, SMEs reviewed these attributes to determine their relative importance to the Air Force cyber career fields.

Literature review

The literature review included select published articles, papers, technical reports, military publications and doctrine, as well as briefings made available to the project team by the USAF and other military sources. First, the Occupational Information Network (O*NET) was used to draft a starting list of important aptitudes and traits for cyber career fields in the Air Force. Developed under the sponsorship of the US Department of Labor/Employment and Training Administration (USDOL/ETA), the O*NET (https://www.onetonline.org/) contains information on hundreds of occupations and describes working in the United States today. O*NET provides essential information for each occupation, including coverage for the skills and knowledge required of job incumbents; as well as the abilities, interests, and work styles found in those same individuals.

One useful feature of O*NET is called Crosswalk. This feature allows Veterans to identify occupations that involve similar tasks and knowledge, skills, abilities, and other characteristics (KSAOs) as those used in their military occupation. Employing Crosswalk, 20 civilian occupations that are similar to the Air Force Specialty Codes (AFSCs) that are the focus of this work were identified. A web scraping of those 20 occupations was performed using R software to automate the extraction of relevant data (e.g., importance ratings, tasks, abilities, work styles) from O*NET into a .csv file. This involved setting up script in R to search the crosswalk for each of the six AFSCs ("3D0X2," "3D0X3," "1B4X1," "1N4X1A," "17DX," "17Sx") and scraping the relevant O*NET data from each of the 20 corresponding civilian jobs. The web scraping resulted in the compilation of a complete list of tasks and KSAOs for civilian occupations that are similar to the six Air Force occupations that are the focus of this work.

To identify important constructs for the cyber AFSCs, both the abilities and work styles identified as essential for the related civilian occupations were included (see Tables 1 and 2). Research supports that "abilities are relatively enduring basic capacities for performing a wide range of different tasks" (Carroll, 1993; Fleishman & Reilly, 1992; Peterson et al., 2001, p. 457). Although extremely useful, abilities do not capture the whole story on worker characteristics or requirements. As such, work styles, which are defined on O*NET as "personal characteristics that can affect how well someone performs a job," were also included in the draft framework.

To help determine the relative importance of these abilities, a relative criticality score was computed. Each ability has two numbers associated with it, one indicating the number of cyber occupations in O*NET where the ability is found; and the second indicating the mean importance of that ability to the occupations (these are reported in Table 1). A mean criticality weight was computed for each type of ability by multiplying the number of occupations times the importance weight from O*NET. This provided a relative weighting as a function of the number of cyber occupations in which it is found and the importance weight for that ability as computed in O*NET.

As with abilities, web scraping was used to identify the work styles associated with the 20 civilian cyber occupations identified via the O*NET crosswalk feature. Table 2 reports the work styles, number of occupations each style is descriptive of, the mean importance rating (range is 0–100), and a criticality score.

Examination of Table 2 reveals that 16 work styles are critical to cyber occupations. Following the procedure described for aptitudes, a relative criticality score was computed for each work style. The criticality scores indicated that Attention to Detail, Dependability, Integrity, and Analytical Thinking were the most influential work styles.

O*NET was a rich source of information for cyber occupations. One particular strength was that it cuts across all cyber occupations in the US economy, those in both the public and private sectors. In all, 19 separate abilities and 16 specific work styles were deemed important to incumbents in cyber occupations.

Next, papers, technical reports, military publications and doctrine, and briefings made available to the project team by the USAF and other military sources were reviewed to identify any additional constructs to be included. A previous stream of research, in which the Air Force commissioned the University of Maryland Center for Advanced Study of Language (CASL) to develop the Cyber Aptitude and Talent Assessment (AF-CATA) battery provided an additional 14 constructs for inclusion in the framework (O'Rourke et al., 2017).

The archival materials were reviewed to examine the extent to which the framework of abilities and traits matched the requirements for the Air Force career fields of interest. From the details in the archival materials, tasks and KSAOs were matched into the constructs identified. These occupational details were drawn primarily from the Air Force Enlisted Classification Directory

Table 1. O*NET abilities for civilian cyber occupations.

Ability	Definition (The ability to . . .)	# POS	IMP	CRIT
Written Comprehension	Read and understand information and ideas presented in writing.	19	73.24	1391
Oral Comprehension	Listen to and understand information and ideas presented through spoken words and sentences.	19	73.12	1389
Oral Expression	Communicate information and ideas in speaking so others will understand.	19	71.12	1351
Problem Sensitivity	Tell when something is wrong or is likely to go wrong. It does not involve solving the problem, only recognizing there is a problem.	19	70.94	1348
Deductive Reasoning	Apply general rules to specific problems to produce answers that make sense.	19	70.53	1340
Inductive Reasoning	Combine pieces of information to form general rules or conclusions (includes finding a relationship among seemingly unrelated events).	19	69.59	1322
Information Ordering	Arrange things or actions in a certain order or pattern according to a specific rule or set of rules (e.g., patterns of numbers, letters, words, pictures, mathematical operations).	19	67.41	1281
Written Expression	Communicate information and ideas in writing so others will understand.	19	67.24	1277
Category Flexibility	Generate or use different sets of rules for combining or grouping things in different ways.	19	56.94	1082
Fluency of Ideas	Come up with a number of ideas about a topic (the number of ideas is important, not their quality, correctness, or creativity).	18	56.59	1019
Flexibility of Closure	Identify or detect a known pattern (a figure, object, word, or sound) that is hidden in other distracting material.	16	52.00	832
Originality	Come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem.	13	52.41	681
Selective Attention	Concentrate on a task over a period of time without being distracted.	12	50.24	603
Mathematical Reasoning	Choose the right mathematical methods or formulas to solve a problem.	9	47.41	427
Perceptual Speed	Quickly and accurately compare similarities and differences among sets of letters, numbers, objects, pictures, or patterns. The things to be compared may be presented at the same time or one after the other. This ability also includes comparing a presented object with a remembered object.	8	45.76	366
Number Facility	Add, subtract, multiply, or divide quickly and correctly.	8	44.82	359
Visualization	Imagine how something will look after it is moved around or when its parts are moved or rearranged.	6	45.35	272
Speed of Closure	Quickly make sense of, combine, and organize information into meaningful patterns.	3	40.88	123
Time Sharing	Shift back and forth between two or more activities or sources of information (such as speech, sounds, touch, or other sources).	1	37.29	37

POS = Number of positions (out of 20). IMP = Mean importance rating (0–100). CRIT (criticality) = #POS * IMP.

(AFECD), Air Force Officer Classification Directory (AFOCD), Career Field Education and Training Plans (CFETPs), and Occupational Analysis Reports (OARs). Some KSAOs from occupational materials did not fit into any of the constructs previously identified, thus the literature was reviewed for applicable constructs. Specifically, knowledge of information systems or information technology or telecommunications was mentioned for each of the AFSCs, so Information and Technology Aptitude (D. M. Trippe et al., 2008) was identified as a potentially important construct to include.

In sum, the review of O*NET data, military research, scientific outlets, and archival military documentation resulted in the identification of 50 aptitudes and traits important for Air Force cyber positions.

Subject matter expert (SME) input

The project team interviewed 11 cyber SMEs to provide further insight into the important aptitudes and traits for the relevant cyber career fields. At least one SME from each AFS was interviewed, with the exception of the 3D0X2 and 3D0X3 career fields. However, a SME who had previously worked in a 3D specialty addressed questions based on their experience while in the 3D0X occupations. Additionally, the SME for the 1N4 specialty polled other instructors teaching cyber courses for their input, thus providing us a wider range of perspectives. Finally, several SMEs had received training from other agencies (e.g., the National Security Agency) and worked in teams comprised of individuals from non-Air Force agencies (those mentioned include NSA, Navy, and Army). Based on the breadth of experience by the collective group of SMEs that were interviewed, the coverage of each AFSC was determined to be sufficient.

Table 3 below summarizes the constructs that SMEs mentioned during the interviews. The source column in the table indicates constructs that overlap with the O*NET or CASL detailed above and the SME generated label means at least one SME provided a different construct that the research described above did not capture. Constructs previously discussed that are omitted from Table 3 were not mentioned when SMEs were asked,

Table 2. O*NET work styles for civilian cyber occupations.

Work Style	Definition (Job Requires . . .)	# POS	IMP	CRIT
Attention to Detail	Being careful about detail and thorough in completing work tasks.	17	89.82	1525
Dependability	Being reliable, responsible, and dependable, and fulfilling obligations.	18	86.53	1555
Integrity	Being honest and ethical.	18	86.41	1549
Analytical Thinking	Analyzing information and using logic to address work-related issues and problems.	18	85.24	1539
Initiative	A willingness to take on responsibilities and challenges.	17	80.41	1363
Cooperation	Being pleasant with others on the job and displaying a good-natured, cooperative attitude.	16	80.24	1281
Adaptability/ Flexibility	Being open to change (positive or negative) and to considerable variety in the workplace.	17	78.94	1341
Persistence	Persistence in the face of obstacles.	15	78.06	1173
Stress Tolerance	Accepting criticism and dealing calmly and effectively with high stress situations.	17	76.24	1293
Achievement/ Effort	Establishing and maintaining personally challenging achievement goals and exerting effort toward mastering tasks.	15	75.24	1131
Self-Control	Maintaining composure, keeping emotions in check, controlling anger, and avoiding aggressive behavior, even in very difficult situations.	15	73.94	1105
Independence	Developing one's own ways of doing things, guiding oneself with little or no supervision, and depending on oneself to get things done.	15	72.24	1086
Innovation	Creativity and alternative thinking to develop new ideas for and answers to work-related problems.	15	68.24	1024
Leadership	Willingness to lead, take charge, and offer opinions and direction.	15	67.24	998
Concern for Others	Being sensitive to others' needs and feelings and being understanding and helpful on the job.	15	62.94	939
Social Orientation	Preferring to work with others rather than alone and being personally connected with others on the job.	15	56.00	830

POS = Number of positions (out of 20). IMP = Mean importance rating (0–100). CRIT (criticality) = #POS * IMP.

"What helps to differentiate between the most successful workers versus average ones in terms of abilities, skills, and traits?"

Consistent with the O*NET data, multiple SMEs identified Oral Expression and Comprehension, Complex Problem Solving, Inductive and Deductive Reasoning, and Written Expression and Comprehension as important abilities for Air Force cyber positions. SME input demonstrated similar consistency with the O*NET and CASL sourced traits, specifically Achievement/Effort, Need for Cognition, Independence, Problem Sensitivity, Conscientiousness, and Stress Tolerance. SMEs also generated several traits and abilities that were not previously identified. The SME-generated traits and aptitudes that were mentioned most frequently were Team Player, Flexibility/Adaptability, Self-Discipline, and Creativity.

Several of the SME-generated constructs overlap with those identified in the scientific literature as important for cyber security professionals. Specifically, Jose et al. (2016) stressed the importance of the abilities related to Working Memory, Cognitive Flexibility, and Systems Thinking. Working memory, or "memorization" as indicated by one SME, is the system that holds multiple pieces of transitory information in the mind, where they can be manipulated (Baddeley & Hitch, 1974)" (Jose et al., 2016, p. 173). Working memory is especially important for cyber-security personnel as they must digest and monitor large quantities of information on a daily basis (Jose et al., 2016). For instance, one of the main responsibilities outlined in the AFECD for 1N4s is to analyze and exploit intelligence information. This involves understanding and acting on large amounts of information. O'Rourke et al. (2017) also recognized Working Memory, specifically visuospatial working memory, as an important aptitude for cyber related AFSs.

Cognitive Flexibility, or "Mental Agility" or "Adaptability/Flexibility," refers to "a person's ability to restructure his or her knowledge as an adaptive response to changing situational demands" (Jose et al., 2016, p. 173). Jose et al. (2016) highlight the importance of this aptitude for cyber personnel because of the demanding activities the occupations require, such as adapting to new technologies that adversaries may be able to exploit. Cognitive Flexibility relates to two abilities previously discussed, Complex Problem Solving and Category Flexibility.

Systems Thinking is another aptitude considered important for cyber occupations by both SME and literature sources (Dawson & Thomson, 2018; Jose et al., 2016). Systems Thinking is defined as "an approach to problem-solving in which an individual possesses an understanding of how multiple parts of a system interact and influence each other (Aronson, 1996)" (Jose et al., 2016, pp. 173–174). Similar to Complex Problem Solving and Information Ordering, Systems Thinking is important for understanding the interactions between various system components, such as a 1N4s requirement to gain and maintain knowledge of intelligence organizations and systems (*Air Force Enlisted Classification Directory (AFECD), 2017*). Further, there is ongoing

Table 3. SME generated constructs.

Construct	Source	# of SMEs who endorsed
Achievement/Effort	O*NET	7
Team Player	SME Generated	7
Flexibility/Adaptability	SME Generated	4
Need for Cognition	CASL	4
Self-Discipline	SME Generated	4
Autonomous/Independence	O*NET	3
Creativity	SME Generated	3
Oral Expression	O*NET	3
Problem Sensitivity	O*NET	3
Complex Problem Solving	CASL	2
Conscientiousness	O*NET	2
Deductive Reasoning	O*NET	2
Inductive Reasoning	O*NET	2
Instructing/Teaching	SME Generated	2
Openness to experience	SME Generated	2
Oral Comprehension	O*NET	2
Resilience	SME Generated	2
Spatial Reasoning	SME Generated	2
Stress Tolerance	O*NET	2
Tolerance for Ambiguity	SME Generated	2
Written Comprehension	O*NET	2
Written Expression	O*NET	2
Initiative	O*NET	1
Active Learning	SME Generated	1
Category Flexibility	O*NET	1
Cynical view of human nature	SME Generated	1
Decision Making	SME Generated	1
Emotional Intelligence	SME Generated	1
Flexibility of Closure	O*NET	1
Integrity	O*NET	1
Leadership	O*NET	1
Mathematical Reasoning	O*NET	1
Mental Agility	SME Generated	1
Multitasking	SME Generated	1
Number Facility	O*NET	1
Selective Attention	O*NET	1
Situational Awareness	SME Generated	1
Skepticism	SME Generated	1
Speed of Closure	O*NET	1
Systems Thinking	SME Generated	1
Time management	SME Generated	1
Tolerance	SME Generated	1
Visualization	O*NET	1
Confident	SME Generated	1
Memorization	O*NET	1
Mission-oriented	SME Generated	1
Humility	SME Generated	1

work at the Army of Research Institute (ARI) to develop a systems thinking assessment to be used for Army cyber occupations (Adis et al., 2017).

As for traits, several SMEs endorsed personal characteristics, such as "active learning," "thinking outside the box," and "creative" as important for cyber positions. Reinforcing this concept, Jose et al. (2016) identified openness/intellectance as important for cyber professionals and Dawson and Thomson (2018) identified continued learning as essential. It is imperative that cyber personnel are able to develop solutions to emerging threats and continually learn more about their field. Multiple SMEs emphasized that the people who excel in the cyber career fields are those who are passionate about learning more; they "eat, sleep, and breathe this work" and "go home at night and program."

Ultimately, the O*NET data, literature review, and SME input yielded 62 distinct constructs important for cyber aptitude.

Gap analysis

Next, the project team determined where gaps exit in current assessment methods. That is, where no current DoD measures mapped to a particular construct that was identified as important. To accomplish this, two advanced graduate students in industrial-organizational psychology reviewed and sorted the DoD measures into distinct constructs. All discrepancies were discussed to consensus. The review of DoD measures identified 102 constructs measured in the current DoD tests: 9 aptitudes in the Armed Services Vocational Aptitude Battery (ASVAB), 26 personality facets in Tailored Adaptive Personality Assessment System (TAPAS), 10 aptitudes and 6 traits in the Air Force Officer Qualifying Test (AFOQT) Situational judgment Test, 30 personality facets in the Self-Description Inventory (SDI+), 4 aptitudes in the Electronic Data Processing Test (EDPT), 12 aptitudes and 3 traits in the CATA battery, 1 aptitude in the Cyber Test, and 1 aptitude in the Multitasking Test.

Most of the important constructs identified that are related to personality traits are currently covered by an existing DoD measure (e.g., TAPAS, SDI+). However, there are clear gaps for Analytical Thinking, Persistence, Resilience, and Skepticism. In addition to some clear gaps where no related DoD measure was found, there are several instances where the coverage is otherwise lacking. For instance, Integrity is only measured by TAPAS facets, which is a measure that is currently only used for enlisted airmen. Thus, this measure would not necessarily be available to use as part of a selection assessment for the officer cyber positions, though the SDI could be modified to include it. Moreover, there are several instances where the potentially relevant measure is the CATA. However, this measure is still undergoing validation, so the feasibility of using it in a selection context is unknown. Table 4 lists the constructs where there are potential gaps due to limited coverage with DoD measures.

SME survey input

A survey was developed to seek input from those in cyber career fields on the constructs listed in Table 4 and their importance for training, on-the-job performance, and incremental benefit. Career field managers (CFMs) for the AFSCs of interest were provided: 1) a roster of 200 potential respondents for the survey (200 e-mails for each 1N4X1A, 3D1X2, and 17DX/

SX), 2) an e-mail draft and survey link to send to the potential respondents to request they complete the survey, and 3) a follow up e-mail to be sent approximately two weeks later requesting the survey be completed.

The survey asked respondents to rank order traits or aptitudes by importance for three different purposes: 1) the expected incremental benefit provided when measures of the construct are used with existing DoD tests for the selection of cyber operators, 2) the importance of the construct to on-the-job performance, and 3) the importance of the construct to performance in training.

Between distribution of the survey links on February 21, 2019 and April 22, 2019, 306 respondents from the distribution lists clicked on the link, with 290 completing at least one item on the survey (36.3% return rate). The results are summarized in Tables 5 and 6. It is important to note the numbers are meaningful within construct type (trait vs. aptitude) as there are five traits to rank order and only four aptitudes. That is, the range of possible values is different for the traits (1–5) and the aptitudes (1–4).

Examination of Table 5 reveals consistency in the top two traits across the three purposes, with Analytical Thinking and Adaptability rated first and second, respectively. Situational Awareness is rated third for two purposes and fourth for the third, giving it an average rating of 3.33 making it the third most important trait. Persistence and Dependability receive averages of 4.00 and 4.67 across purpose ratings making them the fourth and fifth, respectively, most important traits.

Table 6 presents the ratings across purposes and the average for the aptitudes. The incremental benefit rank order is the same as the average across purpose, (highest to lowest) Systems Thinking, Deductive Reasoning, Active Learning, and Decision Making.

SMEs made one rating where they ignored the distinction of trait versus aptitude and the distinction between the different purposes (i.e., incremental benefit, training, and on-the-job performance). These ratings are provided in Table 7.

Examination of the table reveals a trait (Analytical Thinking) is rated highest, followed by three aptitudes (Deductive Reasoning, Active Learning, and Systems Thinking). These aptitudes are essentially interchangeable in importance level given the closeness of their ratings. Next comes a trait (Adaptability), another aptitude (Decision Making), and finally three traits (Situational Awareness, Persistence and Dependability).

The AFPC requested six aptitudes and traits be identified for inclusion in the serious game. Using the results from the SME data collected, the list of critical aptitudes and traits was refined to Analytical Thinking, Deductive Reasoning, Systems Thinking, Active Learning, Adaptability, and Situational Awareness. The decision was made to generally keep the rank orders of the aptitudes and traits from the survey data, but make a minor switch moving the trait Situational Awareness above the aptitude Decision Making. This was done to maximize the expected incremental variance of the serious game measures by having another trait (rather than aptitude) measured by the assessment game. There are many reliable aptitude measures available in existing DoD tests and there tend to be generally high intercorrelations among aptitude tests (e.g., Ree & Earles, 1990). Therefore, it is expected that trait measures reliably

Table 4. Key gaps: constructs without coverage by DoD measure.

Construct	Related Existing DoD Measure(s)	Alternate Predictor(s)
Active Learning	TAPAS Intellectual Efficiency	None
Decision Making	None	None
Complex problem-solving	CATA (Complex problem-solving)	None
Information Foraging	Cyber Test	None
Problem Sensitivity	None	None
Systems Thinking	None	None
Fluency of Ideas	SDI+ (Imagination)	None
Category Flexibility	None	None
Convergent Creative Thinking	CATA (Convergent creative thinking)	None
Mental Model Ability	CATA (Mental model ability)	None
Deductive Reasoning	None	None
Inductive Reasoning	CATA (Rule induction)	None
Information Ordering	None	None
Mental Agility	None	None
Flexibility of closure	CATA (Resistance to interference)	None
Psychomotor Speed	CATA (Psychomotor speed)	None
Anomaly detection	CATA (Anomaly detection)	None
Resistance to Interference	CATA (Resistance to interference)	None
Speed of Closure	None	None
Visuospatial working memory	CATA (Visuospatial working memory), Air Force Multitasking Test	None
Vigilance	CATA (Vigilance)	None
Oral Comprehension	Air Force Multitasking Test	None
Oral Expression	None	None
Written Expression	None	None
Instructing	None	None
Emotional Intelligence	None	None
Adaptability	TAPAS (Tolerance)	None
Analytical Thinking	None	None
Concern for Others	TAPAS (Consideration)	None
Dependability	TAPAS (Responsibility)	None
Independence	SDI+ (Independent)	None
Initiative	TAPAS (Courage)	TAPAS (Achievement)
Integrity	TAPAS (Virtue)	TAPAS Non-Delinquency
Need for Cognition	CATA (Need for Cognition)	TAPAS (Curiosity)
Need for Cognitive Closure	CATA (Need for Cognitive Closure)	None
Persistence	None	None
Resilience	None	None
Situational Awareness	TAPAS (Situational Awareness)	TAPAS (Adjustment)
Skepticism	None	None

Table 5. Rankings made within construct type (1–5 traits).

Trait	Incremental Benefit (N = 275)	On-the-Job (N = 285)	Training (N = 290)	Average Rank Order across the Three Purposes
Analytical Thinking	4.00 (1)	3.63 (1)	3.88 (1)	3.84 (1)
Adaptability	3.07 (2)	3.19 (2)	3.10 (2)	3.11 (2)
Situational Awareness	2.98 (3)	2.95 (3)	2.68 (4)	2.87 (3.33)
Persistence	2.65 (4)	2.63 (5)	2.92 (3)	2.74 (4)
Dependability	2.30 (5)	2.60 (4)	2.41 (5)	2.44 (4.67)

Rank ordering of importance to: training (1 = average; 3 = high; 5 = extreme) and on-the-job performance (1 = average; 3 = high; 5 = extreme) as well as incremental benefit (1 = least; 5 = highest). Note: number in parentheses indicated within-purpose rank order of importance (1 = most important; 5 = least important).

Table 6. Rankings made within construct type (1–4 aptitudes).

Aptitude	Incremental Benefit (N = 277)	On-the-Job (N = 282)	Training (N = 291)	Average Rank Order across the Three Purposes
Systems Thinking	2.72 (1)	2.61 (1)	2.59 (2)	2.65 (1.33)
Deductive Reasoning	2.69 (2)	2.60 (2)	2.51 (3)	2.60 (2.33)
Active Learning	2.39 (3)	2.28 (4)	2.99 (1)	2.56 (2.67)
Decision Making	2.21 (4)	2.50 (3)	1.91 (4)	2.20 (3.67)

Rank ordering of importance to: training (1 = average; 4 = extreme) and on-the-job performance (1 = average; 4 = extreme) as well as incremental benefit (1 = least; 4 = highest). Note: number in parentheses indicated within-purpose rank order of importance (1 = most important; 4 = least important).

Table 7. Global ratings.

Construct	Global Rating (N = 288)
Analytical Thinking	6.23
Deductive Reasoning	5.37
Active Learning	5.26
Systems Thinking	5.26
Adaptability	5.16
Decision Making	4.74
Situational Awareness	4.52
Persistence	4.41
Dependability	4.06

Higher number indicates greater importance. Ratings made ignoring distinction of trait or aptitude (e.g., against all 9 where 1 = least important and 9 = most important).

assessed through serious games will provide a stronger potential for demonstrating significant improvement in predictive validity/classification efficiency when used with existing DoD tests.

Game development

The second phase of this research focused on the development and pilot testing of a prototype serious cyber game to measure the six aptitudes and traits identified as

Figure 2. Screenshot of main game environment.

critical in the first phase of the research (Adaptability, Analytical Thinking, Deductive Reasoning, Persistence, Situational Awareness, and Systems Thinking). The prototype serious cyber game, Virus Slayer, measures these aptitudes and traits through a fictional setting where the examinee is tasked with managing a viral outbreak.[2] During the main gameplay phases, the examinee is in charge of administering resources (e.g., antivirals, food, water) using different modes of transport (i.e., trucks, helicopters) to clinics based on their status throughout the phases of the outbreak. (see Figure 2 for a snapshot of the game environment).

Gameplay is divided into seven main parts: 1) introduction: the player watches a short introduction and orientation that explains the game context, 2) training: immersive training experience where the examinee can practice and demonstrate comprehension of usage of the game controls, 3) selection mini-game: a mini-game where the examinee is tasked with completing an entrance exam to qualify for managing the outbreak, 4) phase one of gameplay: an initial phase where the examinee is introduced to managing the outbreak and then begins working, 5) phase two of gameplay: an inclement weather phase where there is decreased visibility, 6) deductive reasoning mini-game: examinees are tasked with repairing logic boards that were damaged due to inclement weather, and 7) phase three of gameplay: a virus mutation phase where antivirals are used at twice the previous rate.

Analytical thinking

Analytical Thinking is measured through several virus-related and logistics-related items as part of a selection test administered under the guise of determining whether the examinee qualifies to manage the outbreak. These multiple-choice items ask the examinee to analyze information (e.g., performing calculations, interpreting figures) based on information from the game context. For example, one item asks the examinee to indicate how many hours it takes to load a helicopter given premises that stipulate how many helicopters and trucks can be moved in a certain amount of time. Regardless of performance on the Analytical Thinking items, all examinees are told they pass the test and that they will be in charge of directing the emergency response.

Active learning

Active Learning is measured through several need for cognition trait-based items under the pretense that they need to collect information from the examinee that will help identify people in the future that are effective at outbreak management. The active learning items were modified from Cacioppo and Petty's (1982) need for cognition scale and ask about the examinee's preferences for intellectual challenge versus simple tasks. These items are scored with a 5-point Likert scale (strongly agree to strongly disagree).

Deductive reasoning

The Deductive Reasoning mini-game captures deductive reasoning through a series of questions regarding repairing logic boards under the guise of a lightning strike destroying an electrical circuit board that must be repaired before the candidate can progress in the game. Examinees utilize drag-and-drop functionality to place operators (AND, OR, NOT) and outcomes (TRUE, FALSE) to complete the circuits.

Items from the three mini-game measures (analytical thinking, active learning, and deductive reasoning) were piloted on Amazon's Mechanical Turk (MTurk) to select items for inclusion in the prototype game. Items were dropped or revised based on psychometric properties. Based on the pilot results, 13 items were retained for Analytical Thinking, 11 for Deductive Reasoning, and 8 for Active Learning.

Systems thinking

Systems Thinking consists of performance-based measures captured throughout the three gameplay phases. Specifically, Systems Thinking was assessed using the degree to which the examinee identified an optimal strategy to accomplish core game goals, such as delivering supplies to hospitals in need (i.e., antivirals, food, water) and minimizing infections and deaths. For example, one indicator was efficiency with loading vehicles. Specifically, the extent to which the examinee used all space available on the helicopters or trucks they loaded.

Adaptability

Like Systems Thinking, Adaptability consists of performance-based indicators related to accomplishing core game goals. Adaptability was scored according to how well examinees maintained their performance after new obstacles (i.e., inclement weather, virus mutation) came into play. For instance, with all other indicators held equal, an examinee who's efficiency with loading vehicles went down between phase 1 and phase 2 of gameplay, would have a lower adaptability score than an examinee who's efficiency with loading vehicles remained stable regardless of the new obstacles introduced in each phase of gameplay.

Situational awareness

Situational awareness assessed candidates' ability to attend to competing demands of the game without overlooking details. This was measured using behavioral indicators based on the time clinics spend not being viewed/selected (responsiveness behavior), and the rate at which the examinee changed clinic selections (exploration behavior). As part of the Situational Awareness measure, examinees were also asked to provide six mid-game status checks detailing facts about the current situation (e.g., how many clinics have insufficient antivirals).

Psychometric evaluation

Pilot testing

This section provides a brief overview of the pilot testing and initial psychometric results for the prototype serious game. For more information about development of the measures and the initial psychometrics, see Coovert et al. (2020) and Wiernik et al. (in press).

Participants in the study included 267 MTurk workers and 10 university students. The samples were recruited to reflect the military population of interest by requiring the following criteria: US citizenship, between 18 and 24 years of age, and at least a high school level of education. Participants were provided a link to the full prototype gamified assessment, which encompassed all six measures.

To compute more precise IRT estimates, the pilot sample was included in the analyses for the three mini-game assessments (Analytical Thinking, Active Learning, Deductive Reasoning). The sample characteristics are summarized in Tables 8, 9, 10.

Reliabilities for the mini-game measures (Analytical Thinking, Active Learning, and Deductive Reasoning) were estimated with Cronbach's alpha, IRT marginal reliability, and IRT empirical reliability. IRT marginal reliability was calculated using the test information function (based on the item parameters) and assuming a standard normal score distribution using the following formula:

$$r_{xx'} = \int_{-\infty}^{\infty} \frac{I(z)}{I(z)+1}\phi(z)dz$$

where $I(z)$ is the test information at trait level z and $\phi(z)$ is the density of the standard normal distribution at trait level z. Empirical reliability was computed using the estimated factor scores and their associated standard errors:

$$r_{xx'} = \mathrm{Var}(\hat{z}_i)/\left(\mathrm{Var}(\hat{z}_i) + \overline{SE^2}_{\hat{z}_i}\right)$$

The IRT reliabilities were estimated using a 2-parameter logistic (2PL) model. For the Analytical Thinking scale, the alpha, IRT marginal reliability, and IRT empirical reliability all reached acceptable levels (.81, .83, .78, respectively). The Active Learning scale reliability estimates also reached acceptable levels (.86, .87, .87). The Deductive Reasoning scale reliability estimates were lower, but acceptable (.74, .73, .75).

Behavioral indicators based on goal-related metrics were used to create a composite measure of Systems Thinking. Estimates of omega, alpha, and the correlation with unit-weighted composite were used to estimate the reliability of the Systems Thinking composite. The Systems Thinking reliability estimates indicated high levels of reliability (ω = .90, α = .89, r = .96).

Adaptability was based on similar behavioral indicators, captured the degree to which examinees adapt to changes (e.g., more clinics to manage, inclement weather, virus mutation). To accomplish this, the game phase 2 and phase 3 indicators were regressed on the phase 1 indicators then the Adaptability indicators were computed as the residuals of that regression model. The Adaptability reliability estimates indicated acceptable levels of reliability (ω = .80, α = .80, r = .91).

The Situational Awareness measure was based on a mixture of behavioral indicators and mid-game status checks. The composite based on the behavioral indicators demonstrated acceptable reliability (ω = .82, α = .79) and the overall unit-weighted composite of the behavioral indicators and estimated factor scores from the status checks also demonstrated acceptable reliability (α = .75).

Measure intercorrelations were examined to determine whether the dimensions could be distinguished from one another. Intercorrelations among the measures provided evidence that the measures were capturing different underlying constructs (see Table 11). The correlations between different measures ranged from .01 (Analytical Thinking and Active Learning) to .52 (Situational Awareness and Systems Thinking).

Overall, the initial psychometric analyses of the pilot data demonstrated that the mini-games and core gameplay-based assessments have good reliability. Additionally, the scores for the six assessed constructs showed discriminant validity, indicating there was minimal overlap between the measures of the theoretically distinct constructs.

Usability analyses

Usability assessment of the game was determined through an adapted version of the System Usability Scale (SUS). The scale originally was developed by Brooke (1986) and

Table 8. Sample descriptive statistics.

Education Level	Sample 1	Sample 2	Total Sample
Less than high school	4	0	4 (0.5%)
High school	49	32	81 (10.9%)
Some college	106	123	229 (30.8%)
Associate's degree	57	14	71 (9.6%)
Bachelor's degree	201	107	308 (41.5%)
Master's degree	42	0	42 (5.7%)
Doctoral degree	1	0	1 (0.1%)
Professional degree	6	0	6 (0.8%)
No response	0	1	1 (0.1%)
Total	**466**	**277**	**743 (100.0%)**
Race	Sample 1	Sample 2	Total Sample
American Indian or Alaska Native	5	2	7 (0.9%)
Asian	23	21	44 (5.9%)
Black or African American	69	29	98 (13.2%)
Mixed Race	14	20	34 (4.6%)
White	347	195	542 (72.9%)
Other	3	7	10 (1.3%)
Prefer not to answer	5	3	8 (1.1%)
Total	**466**	**277**	**743 (100.0%)**
Ethnicity	Sample 1	Sample 2	Total Sample
Hispanic, Latino/a/e, Spanish	8	39	47 (6.3%)
Prefer not to answer	9	5	14 (1.9%)
Not Hispanic, Latino/a/e, or Spanish	449	233	682 (91.8%)
Total	**466**	**277**	**743 (100.0%)**

Table 9. Correlations between measures.

	Measures (factor scores)						
	ST	AD	SA	SA_bh	AT	AL	DR
ST							
AD	0.43						
SA	0.52	0.20					
SA_bh	0.22	0.05	0.80				
AT	0.25	0.06	0.21	0.08			
AL	0.07	0.04	−0.09	−0.11	0.01		
DR	0.16	0.02	0.05	−0.12	0.35	0.19	

ST = Systems Thinking, AD = Adaptability, SA = Situational Awareness, SA_bh = Situational Awareness composite based on behavioral indicators, AL = Active Learning, DR = Deductive Reasoning.

is widely used by businesses and government (https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html). The scale has widespread acceptance for measuring usability of all types (e.g., websites, interfaces, software and applications among others). See the modified 10 items under Usability in the supplemental materials. The scale is primarily descriptive in nature and extreme scores can indicate problematic areas in the game. For the current effort, it is important to remember the game was developed to assess aptitudes and traits, not for entertainment. As such, interpretation of scores need to keep this goal in mind.

The first assessment of usability occurred during development and testing of the three primary game flow segments. That is, players completed the three segments where they battled the virus but played no mini-games. Examination of the usability items indicates nearly all scores are within plus or minus one standard deviation of the mean of the scale (details available in the supplemental materials). This reflects positively as there are no extreme negative or positive reactions. The one exception is the item dealing with the need of a technical person to assist in the play of the game. Individuals indicated a technical person would be helpful. This is likely due to the difficulty of the resource allocations in the game.

The effectiveness of the training was assessed with five items. Examination of these means indicated the players felt prepared to play the game. Help features were assessed with two items and examinee responses indicated the features left the players at least somewhat prepared. The average number of times individuals accessed help during game play was less than two, so this speaks well of the training and structure of the gameplay.

Finally, one item asked for a global impression of the user friendliness of the game. Players indicated overall it is between OK and good; an acceptable level given the purpose of the game. The next point at which usability was assessed was for the deployment of the entire game, with full mini-games and game play modules in place.

Data were gathered and results from a sample of 139 examinees with listwise deletion is available in the supplemental materials. The same items used in the first assessment of usability were used. Comparison of this sample to the previous one revealed some movement in the means, but all within plus or minus one standard deviation.

So descriptively, the usability of the game is positive. Thus, based on the perception of usability from MTurk workers, the serious game can be used as an assessment. It will be important, however, to continue to assess usability when the game is utilized on a sample of Air Force personnel.

Discussion

Recent technological advances have led to a surge in interest surrounding the use of serious games for providing immersive and realistic simulations that may allow for the elicitation and ultimate measurement of constructs that are difficult to measure through traditional assessments (Fetzer, 2015; Landers et al., 2018; Tippins, 2015). Overall, evidence from pilot testing provides preliminary support for the use of a serious game for cyber aptitude assessment.

During Phase I of this effort, archival information was reviewed, and cyber career field managers were interviewed, to identify aptitudes and traits required for success in selected Air Force enlisted and officer cyber careers. Next, Air Force personnel specialists identified which cyber aptitudes and traits could be measured through existing Air Force or DoD tests and where measurement gaps existed. Nine aptitudes and traits were identified as the most closely related to on-the-job cyber performance for which there was inadequate coverage and could be captured by a serious game (Coovert et al., 2019).

At the beginning of Phase II, six aptitudes and traits were targeted for assessment in the serious game (Active Learning, Adaptability, Analytical Thinking, Deductive Reasoning, Situation Awareness, and Systems Thinking). Assessment occurs through a combination of game play and a series of mini-games embedded in the main game. Initial psychometric evaluations were done with nonmilitary samples selected to resemble military applicants in age and education level (Coovert et al., 2020).

Item response theory (IRT) and factor analysis methods were used to examine the construct validity of the indicators used to assess the six constructs. IRT analyses indicated good marginal reliabilities for all six constructs. Indicator variables were developed to assess Adaptability, Situational Awareness, and Systems Thinking from game play. The IRT and factor analyses indicated reasonable construct validity and discrimination for the items on those constructs.

Phase III of this project will involve additional psychometric analyses with US Air Force personnel. It has been delayed due to restrictions on data collection activities during the COVID-19 pandemic. Additional data collection will occur with US Air Force Basic Recruits Lackland AFB when restrictions have been lifted. These recruits also will have taken the ASVAB and TAPAS, which will facilitate evaluation of construct validity of the cyber serious game. The cyber serious game also will be administered to Air Force cyber trainees at Keesler AFB to examine its predictive validity and incremental validity against training performance.

Limitations

The primary limitation of this study is the lack of psychometric data for military samples. Phase III data collection will address this deficiency when Air Force Basic Recruits at Lackland AFB are tested to evaluate the usability of the serious game and its psychometric characteristics including construct validity and subgroup differences. Cyber trainees at Keesler AFB will be tested to evaluate the predictive validity of the game and its incremental validity when used with other measures.

The validation data will help address critical unknowns with regard to the use of serious games for assessment. One critical unknown is whether serious games can produce more valid assessments, either through improved measurement of known constructs, or through measurement of new job-related constructs. A second important unknown is whether serious gaming can produce more efficient assessments (e.g., through measurement of multiple traits simultaneously), potentially reducing administration time, without losing validity achieved through more traditional methods.

Notes

1. The terms traits, personality factors/facets, and work styles are used interchangeably throughout this paper, while aptitude is used interchangeably with ability. These terms are collectively referred to as constructs.
2. The concept and prototype for the serious game was determined, and game development began, approximately one year prior to the identification of COVID-19 in China and the resulting world-wide pandemic.

Data availability statement

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Adis, C., Wisecarver, M., Raber, C., Wind, A. P., & Canali, K. G. (2017). *Innovative tools to assess systems thinking ability*. U.S. Army Research Institute for the Behavioral and Social Sciences.

Air Force Enlisted Classification Directory (AFECD) (p. 361). (2017). Air Force Personnel Center.

Aronson, D. (1996). *Overview of systems thinking.* (Vol. 8, (2009)). http://www.thinking.net/Systems_Thinking/systems_thinking.htmlS

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower, (Ed.), *The Psychology of Learning and Motivation* (pp. 47–89). Academic Press.

Brooke, J. (1986). System usability scale (sus): A quick-and-dirty method of system evaluation user information. *Digital equipment co ltd. Reading, UK*, 1–7.

Bureau of Labor Statistics, U.S. Department of Labor. (2021). *Information security analysts: Occupational outlook handbook: U.S. bureau of labor statistics*. https://www.bls.gov/ooh/computer-and-information-technology/information-security-analysts.htm

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, *42*(1), 116.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge University Press.

Coovert, M. D., Martin, J., Howard, D. J., Kim, F., Dreibelbis, R., Arbogast, M. S., & Potter, S. (2019). *Use of technology-enhanced work simulations for cyber aptitude assessment* (AFRL-RH-WP-TR-2019-0043). Air Force Research Laboratory, 711 Human Performance Wing, Airman Systems Directorate, Collaborative Interfaces and Teaming Branch.

Coovert, M. D., Wiernik, B. M., & Martin, J. (2020). *Use of technology-enhanced simulations for cyber aptitude assessment: Phase II prototype development* (AFRL-RH-WP-TR-2020-0055). Air Force Research Laboratory, 711 Human Performance Wing, Airman Systems Directorate, Collaborative Interfaces and Teaming Branch.

Dawson, J., & Thomson, R. (2018). The future cybersecurity workforce: Going beyond technical skills for successful cyber performance. *Frontiers in Psychology*, *9*, 744. https://doi.org/10.3389/fpsyg.2018.00744

Drasgow, F., Embretson, S. E., Kyllonen, P. C., & Schmitt, N. (2006). *Technical review of the armed services vocational aptitude battery (ASVAB)* (FR No. 06–25). Human Resources Research Organization.

Dreibelbis, R. C., Martin, J., Coovert, M. D., & Dorsey, D. W. (2018). The looming cybersecurity crisis and what it means for the practice of industrial and organizational psychology. *Industrial and Organizational Psychology*, *11*(2), 346–365. https://doi.org/10.1017/iop.2018.3

Exec. Order No. 13870, 3 C.F.R. 20523. (2019). https://www.federalregister.gov/documents/2019/05/09/2019-09750/americas-cybersecurity-workforce

Fetzer, M. (2015). Serious games for talent selection and development. *TIP: The Industrial-Organizational Psychologist*, *52*(3), 117–125.

Fleishman, E. A., & Reilly, M. E. (1992). *Handbook of human abilities: Definitions, measurements, and job task requirements*. Consulting Psychologists Press.

Gould, J. (2013). Cyber soldiers: Army set to load up 2 new enlisted MOSs. *Army Times*, *1*, 8–19.

Handler, C. (2013). Foreward. In M. Fetzer & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. v–ix). Springer.

Jose, I., LaPort, K., & Trippe, D. M. (2016). Requisite attributes for cyber security personnel and teams: Cyber risk mitigation through talent management. In Stephen J. Zaccaro, Reeshad S. Dalal, Lois E. Tetrick, & Julie A. Steinke (Eds.), *Psychosocial dynamics of cyber security* (pp. 167–198). Routledge.

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, Its history and future: Definitions and a research agenda. *Simulation & Gaming*, *49*(3), 315–338. edb. https://doi.org/10.1177/1046878118774385

McMichael, W. H. (2010). *DoD cyber command is officially online. Army Times.* http://www.airforcetimes.com/news/2010/05/military_cyber_command_052110

O'Rourke, P., Karuzis, V. P., Kim, S., Tseng, A., Pandža, N. B., Young, J. M., Colflesh, G. J. H., Hughes, M. M., Liter, A., Lee, J., Harbison, J. I., Sumer, A., & Campbell, S. G. (2017). *Assessing aptitude for USAF cyber warfare operations training.* University of Maryland Center for Advanced Study of Language.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., Campion, M. A., Mayfield, M. S., Morgeson, F. P., & Pearlman, K. (2001). Understanding work using the occupational information network (O* NET): Implications for practice and research. *Personnel Psychology*, *54*(2), 451–492. https://doi.org/10/ds3d77

Ree, M. J., & Earles, J. (1990). *Differential validity of a differential aptitude test* (AFHRL-TR-89-59). Air Force Human Resources Laboratory.

Tippins, N. T. (2011). Overview of technology-enhanced assessments. In N. T. Tippins & S. Adler (Eds.), *Technology-enhanced assessment of talent* (pp. 1–18). John Wiley & Sons.

Tippins, N. T. (2015). Technology and assessment in selection. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 551–582. https://doi.org/10.1146/annurev-orgpsych-031413-091317

Trippe, D. M., Russell, T. L., Schwartz, K. L., & Weissmuller, J. J. (2008). *Issues in information and communication technology test development: A literature review and summary of best practice.* Air Force Personnel Center.

Trippe, M. D., Moriarty, K. O., Russell, T. L., Carretta, T. R., & Beatty, A. S. (2014). Development of a cyber/information technology knowledge test for military enlisted technical training qualification. *Military Psychology*, *26*(3), 182–198. https://doi.org/10/gfn5hp

Wiernik, B. M., Raghavan, M., Coovert, M. D., & Carretta, T. R. (in press). Developing and validating serious game-based assessment for cyber operations in the U.S. air force. *International Journal of Selection and Assessment*.