

Advances in Intelligent Systems and Computing 1271

Kim-Kwang Raymond Choo
Tommy Morris
Gilbert L. Peterson
Eric Imsand *Editors*

National Cyber Summit (NCS) Research Track 2020

 Springer

Advances in Intelligent Systems and Computing

Volume 1271

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

Advisory Editors

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India

Rafael Bello Perez, Faculty of Mathematics, Physics and Computing,
Universidad Central de Las Villas, Santa Clara, Cuba

Emilio S. Corchado, University of Salamanca, Salamanca, Spain

Hani Hagra, School of Computer Science and Electronic Engineering,
University of Essex, Colchester, UK

László T. Kóczy, Department of Automation, Széchenyi István University,
Gyor, Hungary


Vladik Kreinovich, Department of Computer Science, University of Texas
at El Paso, El Paso, TX, USA

Chin-Teng Lin, Department of Electrical Engineering, National Chiao
Tung University, Hsinchu, Taiwan

Jie Lu, Faculty of Engineering and Information Technology,
University of Technology Sydney, Sydney, NSW, Australia

Patricia Melin, Graduate Program of Computer Science, Tijuana Institute
of Technology, Tijuana, Mexico

Nadia Nedjah, Department of Electronics Engineering, University of Rio de Janeiro,
Rio de Janeiro, Brazil

Ngoc Thanh Nguyen , Faculty of Computer Science and Management,
Wrocław University of Technology, Wrocław, Poland

Jun Wang, Department of Mechanical and Automation Engineering,
The Chinese University of Hong Kong, Shatin, Hong Kong

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****


More information about this series at <http://www.springer.com/series/11156>

Kim-Kwang Raymond Choo ·
Tommy Morris · Gilbert L. Peterson ·
Eric Imsand
Editors

National Cyber Summit (NCS) Research Track 2020

 Springer

Editors

Kim-Kwang Raymond Choo 
Department of Information Systems
The University of Texas at San Antonio
San Antonio, TX, USA

Tommy Morris
The University of Alabama in Huntsville
Huntsville, AL, USA

Gilbert L. Peterson
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA

Eric Imsand
The University of Alabama in Huntsville
Huntsville, AL, USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-030-58702-4 ISBN 978-3-030-58703-1 (eBook)
<https://doi.org/10.1007/978-3-030-58703-1>

© Springer Nature Switzerland AG 2021

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Cyberthreats remain important and strategically relevant in both developed and developing economies. For example, in the “Worldwide threat assessment of the US intelligence community (January 29, 2019)¹, it was reported that:

Our adversaries and strategic competitors will increasingly use cyber capabilities – including cyber espionage, attack, and influence—to seek political, economic, and military advantage over the United States and its allies and partners

Hence, there is a need to keep a watchful brief on the cyberthreat landscape, particularly as technology advances and new cyberthreats emerge. This is the intention of this conference proceedings.

This conference proceedings contains 12 regular papers and 4 short papers from the 2020 National Cyber Summit Research Track. The proceedings also includes one invited paper. The 2020 National Cyber Summit was originally planned to be held in Huntsville, Alabama, from June 2 to 4, 2020. However, due to the COVID-19 pandemic, all tracks of the 2020 National Cyber Summit, except the Research Track, were delayed until June 2021. The 2020 National Cyber Summit Research Track was held via video conference on June 3 and 4. Authors from each selected paper presented their work and took questions from the audience.

The papers were selected from submissions from universities, national laboratories, and the private sector from across the USA, Bangladesh, and Slovenia. All of the papers went through an extensive review process by internationally recognized experts in cybersecurity.

The Research Track at the 2020 National Cyber Summit has been made possible by the joint effort of a large number of individuals and organizations worldwide. There is a long list of people who volunteered their time and energy to put together the conference and deserved special thanks. First and foremost, we would like to offer our gratitude to the entire Organizing Committee for guiding the entire process

¹<https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR—SSCI.pdf>.

of the conference. We are also deeply grateful to all the Program Committee members for their time and efforts in reading, commenting, debating, and finally selecting the papers. We also thank all the external reviewers for assisting the Program Committee in their particular areas of expertise as well as all the authors, participants, and session chairs for their valuable contributions.

Tommy Morris
Kim-Kwang Raymond Choo
General Chairs
Gilbert L. Peterson
Eric Imsand
Program Committee Chairs

Organization

Organizing Committee

General Chairs

Tommy Morris The University of Alabama in Huntsville, USA
Kim-Kwang Raymond Choo The University of Texas at San Antonio, USA

Program Committee Chairs

Gilbert L. Peterson Air Force Institute of Technology, USA
Eric Imsand The University of Alabama in Huntsville, USA

Program Committee and External Reviewers

Program Committee Members

Tommy Morris The University of Alabama in Huntsville, USA
Kim-Kwang Raymond Choo The University of Texas at San Antonio, USA
George Grispos The University of Nebraska at Omaha, USA
Eric Imsand The University of Alabama in Huntsville, USA
Nour Moustafa The University of New South Wales at Canberra,
 Australia
Wei Zhang The University of Louisville, USA
Reza M. Parizi Kennesaw State University, USA
Octavio Loyola-González Tecnologico de Monterrey, Mexico
Ezhil Kalaimannan The University of West Florida, USA
Vijayan Sugumaran Oakland University, USA
Vahid Heydari Rowan University, USA
Jianyi Zhang Beijing Electronic Science and Technology
 Institute, China
Gilbert Petersen US Air Force Institute of Technology, USA
Ravi Rao Fairleigh Dickinson University, USA

Jun Dai	California State University, Sacramento, USA
David Dampier	Marshall University, USA
Patrick Pape	Auburn University Montgomery, USA
Cong Pu	Marshall University, USA
Junggab Son	Kennesaw State University, USA
John Bland	The University of Alabama in Huntsville, USA
Robin Verma	The University of Texas at San Antonio, USA
Miguel Angel Medina Pérez	Tecnologico de Monterrey, Mexico
Huijun Wu	Arizona State University, USA
Rongxing Lu	The University of New Brunswick, Canada

External Reviewers

Arquímedes Méndez Molina
 Daniel Colvett
 Duo Lu
 Erdal Kose
 Feng Wang
 Gautam Srivastava
 Hossain Shahriar
 Iman Vakilinia
 Irene Kasian
 James Okolica
 Jorge Rodriguez Ruiz
 Lazaro Bustio
 Lei Xu
 Mariano Vargas-Santiago
 Miguel Ángel Álvarez Carmona
 Mohammad Shojaeshafiei
 Mohiuddin Ahmed
 Nickolaos Koroniotis
 Ning Wang
 Nisheeth Agrawal
 Rachel Foster
 Seyedamin Pouriyeh
 Swapnoneel Roy
 Tania Williams
 Tymaine Whitaker
 Xiaomei Zhang
 Yan Huang
 Yaoqing Liu
 Younghun Chae
 Zachary Tackett
 Zhimin Gao

Contents

Invited Paper

Experiences and Lessons Learned Creating and Validating Concept Inventories for Cybersecurity	3
Alan T. Sherman, Geoffrey L. Herman, Linda Oliva, Peter A. H. Peterson, Enis Golaszewski, Seth Poulsen, Travis Scheponik, and Akshita Gorti	

Cyber Security Education

A Video-Based Cybersecurity Modular Lecture Series for Community College Students	37
Anton Dahbura and Joseph Carrigan	

Digital Forensics Education Modules for Judicial Officials	46
Ragib Hasan, Yuliang Zheng, and Jeffery T. Walker	

Gaming DevSecOps - A Serious Game Pilot Study	61
James S. Okolica, Alan C. Lin, and Gilbert L. Peterson	

A Single-Board Computing Constellation Supporting Integration of Hands-On Cybersecurity Laboratories into Operating Systems Courses	78
Jason Winningham, David Coe, Jeffrey Kulick, Aleksandar Milenkovic, and Letha Etzkorn	

TWOPD: A Novel Approach to Teaching an Introductory Cybersecurity Course	92
Chola Chhetri	

Network System's Vulnerability Quantification Using Multi-layered Fuzzy Logic Based on GQM	100
Mohammad Shojaeshafiei, Letha Etzkorn, and Michael Anderson	

Cyber Security Technology

Challenges of Securing and Defending Unmanned Aerial Vehicles 119
 William Goble, Elizabeth Braddy, Mike Burmester, Daniel Schwartz,
 Ryan Sloan, Demetra Drizis, Nitish Ahir, Melissa Ma, Matt Bays,
 and Matt Chastain

**Using Least-Significant Bit and Random Pixel Encoding
 with Encryption for Image Steganography** 139
 Tapan Soni, Richard Baird, Andrea Lobo, and Vahid Heydari

**Automated Linux Secure Host Baseline for Real-Time Applications
 Requiring SCAP Compliance** 154
 Zack Kirkendoll, Matthew Lueck, Nathan Hutchins, and Loyd Hook

Deficiencies of Compliancy for Data and Storage 170
 Howard B. Goodman and Pam Rowland

**Taming the Digital Bandits: An Analysis of Digital Bank Heists
 and a System for Detecting Fake Messages in Electronic
 Funds Transfer** 193
 Yasser Karim and Ragib Hasan

**Identifying Vulnerabilities in Security and Privacy of Smart
 Home Devices** 211
 Chola Chhetri and Vivian Motti

Short Papers

**Information Warfare and Cyber Education: Issues
 and Recommendations** 235
 Joshua A. Sipper

**Designing an Internet-of-Things Laboratory to Improve Student
 Understanding of Secure Embedded Systems** 238
 A. R. Rao, Kavita Mishra, and Nagasravani Recharla

Distributed Denial of Service Attack Detection 240
 Travis Blue and Hossain Shahriar









Is There a Prophet Who Can Predict Software Vulnerabilities? 242
 Michael T. Shrove and Emil Jovanov

Author Index 245

Invited Paper



Experiences and Lessons Learned Creating and Validating Concept Inventories for Cybersecurity

Alan T. Sherman¹(✉) , Geoffrey L. Herman² , Linda Oliva¹ ,
Peter A. H. Peterson³ , Enis Golaszewski¹ , Seth Poulsen² ,
Travis Scheponik¹ , and Akshita Gorti¹ 

¹ Cyber Defense Lab, University of Maryland, Baltimore County (UMBC),
Baltimore, MD 21250, USA

sherman@umbc.edu

² Computer Science, University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA

glherman@illinois.edu

³ Department of Computer Science, University of Minnesota Duluth,
Duluth, MN 55812, USA

pahp@d.umn.edu

<http://www.csee.umbc.edu/people/faculty/alan-t-sherman/>,

<http://publish.illinois.edu/glherman/>, <http://www.d.umn.edu/~pahp/>

Abstract. We reflect on our ongoing journey in the educational *Cybersecurity Assessment Tools (CATS)* Project to create two concept inventories for cybersecurity. We identify key steps in this journey and important questions we faced. We explain the decisions we made and discuss the consequences of those decisions, highlighting what worked well and what might have gone better.

The CATS Project is creating and validating two concept inventories—conceptual tests of understanding—that can be used to measure the effectiveness of various approaches to teaching and learning cybersecurity. The *Cybersecurity Concept Inventory (CCI)* is for students who have recently completed any first course in cybersecurity; the *Cybersecurity Curriculum Assessment (CCA)* is for students who have recently completed an undergraduate major or track in cybersecurity. Each assessment tool comprises 25 *multiple-choice questions (MCQs)* of various difficulties that target the same five core concepts, but the CCA assumes greater technical background.

Key steps include defining project scope, identifying the core concepts, uncovering student misconceptions, creating scenarios, drafting question stems, developing distractor answer choices, generating educational materials, performing expert reviews, recruiting student subjects, organizing workshops, building community acceptance, forming a team and nurturing collaboration, adopting tools, and obtaining and using funding.

Creating effective MCQs is difficult and time-consuming, and cybersecurity presents special challenges. Because cybersecurity issues are often

subtle, where the adversarial model and details matter greatly, it is challenging to construct MCQs for which there is exactly one best but non-obvious answer. We hope that our experiences and lessons learned may help others create more effective concept inventories and assessments in STEM.

Keywords: Computer science education · Concept inventories · Cryptography · Cybersecurity Assessment Tools (CATS) · Cybersecurity education · Cybersecurity Concept Inventory (CCI) · Cybersecurity Curriculum Assessment (CCA) · Multiple-choice questions

1 Introduction

When we started the *Cybersecurity Assessment Tools (CATS)* Project [44] in 2014, we thought that it should not be difficult to create a collection of 25 *multiple choice questions (MCQs)* that assess student understanding of core cybersecurity concepts. Six years later, in the middle of validating the two draft assessments we have produced, we now have a much greater appreciation for the significant difficulty of creating and validating effective and well-adopted concept inventories. This paper highlights and reflects on critical steps in our journey, with the hope that our experiences can provide useful lessons learned to anyone who wishes to create a cybersecurity concept inventory, any assessment in cybersecurity, or any assessment in STEM.¹

Cybersecurity is a vital area of growing importance for national competitiveness, and there is a significant need for cybersecurity professionals [4]. The number of cybersecurity programs at colleges, universities, and training centers is increasing. As educators wrestle with this demand, there is a corresponding awareness that we lack a rigorous research base that informs how to prepare cybersecurity professionals. Existing certification exams, such as CISSP [9], are largely informational, not conceptual. We are not aware of any scientific analysis of any of these exams. Validated assessment tools are essential so that cybersecurity educators have trusted methods for discerning whether efforts to improve student preparation are successful [32]. The CATS Project provides rigorous evidence-based instruments for assessing and evaluating educational practices; in particular, they will help assess approaches to teaching and learning cybersecurity such as traditional lecture, case study, hands-on lab exercises, interactive simulation, competition, and gaming.

We have produced two draft assessments, each comprising 25 MCQs. The *Cybersecurity Concept Inventory (CCI)* measures how well students understand core concepts in cybersecurity after a first course in the field. The *Cybersecurity Curriculum Assessment (CCA)* measures how well students understand core concepts after completing a full cybersecurity curriculum. Each test item comprises a *scenario*, a *stem* (a question), and five *alternatives* (answer choices comprising

¹ Science, Technology, Engineering, and Mathematics (STEM).

a single best answer choice and four distractors). The CCI and CCA target the same five core concepts (each being an aspect of adversarial thinking), but the CCA assumes greater technical background. In each assessment, there are five test items of various difficulties for each of the five core concepts.

Since fall 2014, we have been following prescriptions of the National Research Council for developing rigorous and valid assessment tools [26,35]. We carried out two surveys using the Delphi method to identify the scope and content of the assessments [33]. Following guidelines proposed by Ericsson and Simon [13], we then carried out qualitative interviews [40,45] to develop a cognitive theory that can guide the construction of assessment questions. Based on these interviews, we have developed a preliminary battery of over 30 test items for each assessment. Each test item measures how well students understand core concepts as identified by our Delphi studies. The distractors (incorrect answers) for each test item are based in part on student misconceptions observed during the interviews. We are now validating the CCI and CCA using small-scale pilot testing, cognitive interviews, expert review, and large-scale psychometric testing [31].

The main contributions of this paper are lessons learned from our experiences with the CATS Project. These lessons include strategies for developing effective scenarios, stems, and distractors, recruiting subjects for psychometric testing, and building and nurturing effective collaborations. We offer these lessons, not with the intent of prescribing advice for all, but with the hope that others may benefit through understanding and learning from our experiences. This paper aims to be the paper we wished we could have read before starting our project.

2 Background and Previous and Related Work

We briefly review relevant background on concept inventories, cybersecurity assessments, and other related work. To our knowledge, our CCI and CCA are the first concept inventories for cybersecurity, and there is no previous paper that presents lessons learned creating and validating any concept inventory.

2.1 Concept Inventories

A *concept inventory (CI)* is an assessment (typically multiple-choice) that measures how well student conceptual knowledge aligns with accepted conceptual knowledge [23]. Concept inventories have been developed for many STEM disciplines, consistently revealing that students who succeed on traditional classroom assessments struggle to answer deeper conceptual questions [6,14,19,23,28]. When students have accurate, deep conceptual knowledge, they can learn more efficiently, and they can transfer their knowledge across contexts [28]. CIs have provided critical evidence supporting the adoption of active learning and other evidence-based practices [14,19,20,29]. For example, the *Force Concept Inventory (FCI)* by Hestenes et al. [23] “spawned a dramatic movement of reform in physics education.” [12, p. 1018].

For CIs to be effective, they need to be validated. Unfortunately, few CIs have undergone rigorous validation [34,47]. Validation is a chain of evidence that supports the claims that an assessment measures the attributes that it claims to measure. This process requires careful selection of what knowledge should be measured, carefully constructing questions that are broadly accepted as measuring that knowledge, and providing statistical evidence that the assessment is internally consistent. The usefulness of a CI is threatened if it fails any of these requirements. Additionally, a CI must be easy to administer, and its results must be easy to interpret—or they can easily be misused. Critically, CIs are intended as research instruments that help instructors make evidence-based decisions to improve their teaching and generally should not be used primarily to assign student grades or to evaluate a teacher’s effectiveness.

Few validated CIs have been developed for computing topics; notable exceptions include the Digital Logic Concept Inventory [22] (led by CATS team member Herman) and early work on the Basic Data Structures Inventory [37]. None has been developed for security related topics.

2.2 Cybersecurity Assessment Exams

We are not aware of any other group that is developing an educational assessment tool for cybersecurity. There are several existing certification exams, including ones listed by NICCS as relevant [11].

CASP+ [8] comprises multiple-choice and performance tasks items including enterprise security, risk management, and incident response. OSCP [41] (offensive security) is a 24-hour practical test focusing on penetration testing. Other exams include CISSP, Security+, and CEH [7,9,46], which are mostly informational, not conceptual. Global Information Assurance Certification (GIAC) [10] offers a variety of vendor-neutral MCQ certification exams linked to SANS courses; for each exam type, the gold level requires a research paper. We are unaware of any scientific study that characterizes the properties of any of these tests.

2.3 Other Related Work

The 2013 IEEE/ACM Computing Curriculum Review [25] approached the analysis of cybersecurity content in undergraduate education from the perspective of traditional university curriculum development. Later, the ACM/IEEE/AIS SIGSEC/IFIP Joint Task Force on Cybersecurity Education (JTF) [15] developed comprehensive curricular guidance in cybersecurity education, releasing Version 1.0 of their guidelines at the end of 2017.

To improve cybersecurity education and research, the National Security Agency (NSA) and Department of Homeland Security (DHS) jointly sponsor the National Centers of Academic Excellence (CAE) program. Since 1998, more than 300 schools have been designated as CAEs in Cyber Defense. The requirements include sufficiently covering certain “Knowledge Units” (KUs) in their academic programs, making the CAE program a “significant influence on the curricula of programs offering cybersecurity education” [16].

The NICE Cybersecurity Workforce Framework [30] establishes a common lexicon for explaining a structured description of professional cybersecurity positions in the workforce with detailed documentation of the knowledge, skills, and abilities needed for various types of cybersecurity activities.

More recently, the Accreditation Board of Engineering and Technology (ABET) has included, in the 2019–2020 Criteria for Accrediting Computing Program, criteria for undergraduate cybersecurity (or similarly named) programs. ABET has taken an approach similar to that of the CAE program, requiring coverage of a set of topics without requiring any specific set of courses.

In a separate project, CATS team member Peterson and his students [24,36] worked with experts to identify specific and persistent commonsense misconceptions in cybersecurity, such as that “physical security is not important,” or that “encryption is a foolproof security solution.” They are developing a CI focusing on those misconceptions.

3 Key Steps and Takeaways from the CATS Project

We identify the key steps in our journey creating and validating the CCI and CCA. For each step, we comment on important issues, decisions we made, the consequences of those decisions, and the lessons we learned.

3.1 Genesis of the CATS Project

On February 24–25, 2014, Sherman, an expert in cybersecurity, participated in a NSF workshop to advise NSF on how to advance cybersecurity education. NSF occasionally holds such workshops in various areas and distributes their reports, which can be very useful in choosing research projects. The workshop produced a list of prioritized recommendations, beginning with the creation of a concept inventory [5]. At the workshop, Sherman met one of his former MIT officemates, Michael Loui. Sherman proposed to Loui that they work together to create such a concept inventory. About to retire, Loui declined, and introduced Sherman to Loui’s recent PhD graduate Herman, an expert in engineering education. Without meeting in person for over a year, Sherman and Herman began a productive collaboration. Loui’s introduction helped establish initial mutual trust between Sherman and Herman.

3.2 Defining Scope of Project and Assessment Tools

As with many projects, defining scope was one of the most critical decisions of the CATS Project. We pondered the following questions, each of whose answers had profound implications on the direction and difficulty of the project. How many assessment tools should we develop? For what purposes and subject populations should they be developed? In what domain should the test items be cast? Should the test items be MCQs?

We decided on creating two tools: the CCI (for students in any first course in cybersecurity) and CCA (for recent graduates of a major or track in cybersecurity), because there is a strong need for each, and each tool has different requirements. This decision doubled our work. Creating any more tools would have been too much work.

Our driving purpose is to measure the effectiveness of various approaches to teaching cybersecurity, not to evaluate the student or the instructor. This purpose removes the need for a high-stakes test that would require substantial security and new questions for each test instance. By contrast, many employers who have talked with us about our work have stated their desire for an instrument that would help them select whom to hire (our assessments are neither designed nor validated for that high-stakes purpose).

Ultimately, we decided that all test items should be cast in the domain of cybersystems, on the grounds that cybersecurity takes place in the context of such systems. Initially, however, we experimented with developing test items that probed security concepts more generally, setting them in a variety of every-day contexts, such as building security, transportation security, and physical mail. Both approaches have merit but serve different purposes.

Following the format of most concept inventories, we decided that each test item be a MCQ. For more about MCQs and our reasons for using them, see Sect. 4.

3.3 Identifying Core Concepts

The first major step in creating any concept inventory is to identify the core concepts to be tested. We sought about five important, difficult, timeless, cross-cutting concepts. These concepts do not have to cover cybersecurity comprehensively. For example, the Force Concept Inventory targets five concepts from Newtonian dynamics, not all concepts from physics. To this end, we engaged 36 cybersecurity experts in two Delphi processes, one for the CCI and one for the CCA [33]. A Delphi process is a structured process for achieving consensus on contentious issues [3, 18].

An alternative to the Delphi process is the focus group. Although focus groups can stimulate discussions, they can be influenced strongly by personalities and it can be difficult to organize the results coherently. For example, attempts to create concept maps for cybersecurity via focus groups have struggled to find useful meaning in the resulting complex maps, due to their high density.²

Delphi processes also have their challenges, including recruiting and retaining experts, keeping the experts focused on the mission, and processing expert comments, including appropriately grouping similar comments. We started with 36 experts in total, 33 for CCI, 31 for CCA, and 29 in both. We communicated with the experts via email and SurveyMonkey. For each process, approximately 20 experts sustained their efforts throughout. Many of the experts came with

² Personal correspondence with Melissa Dark (Purdue).

strongly held opinions to include their favorite topics, such as policy, forensics, malware analysis, and economic and legal aspects. We completed the two Delphi processes in parallel in fall 2014, taking about eight weeks, conducting initial topic identification followed by three rounds of topic ratings. Graduate research assistant Parekh helped orchestrate the processes. It is difficult to recruit and retain experts, and it is a lot of work to process the large volume of free-form comments.

The first round produced very similar results for both Delphi processes, with both groups strongly identifying aspects of adversarial thinking. Therefore, we restarted the CCI process with an explicit focus on adversarial thinking. After each round, using principles of grounded theory [17], we grouped similar responses and asked each expert to rate each response on a scale from one to ten for importance and timeliness. We also encouraged experts to explain their ratings. We communicated these ratings and comments (without attribution) to everyone. The CCA process produced a long list of topics, with the highest-rated ones embodying aspects of adversarial thinking.

In the end, the experts came to a consensus on five important core concepts, which deal with adversarial reasoning (see Table 1). We decided that each of the two assessment tools would target these same five concepts, but assume different levels of technical depth.

Table 1. The five core concepts underlying the CCI and CCA embody aspects of adversarial thinking.

- 1 Identify vulnerabilities and failures
- 2 Identify attacks against CIA triad^a and authentication
- 3 Devise a defense
- 4 Identify the security goals
- 5 Identify potential targets and attackers

^aCIA Triad (Confidentiality, Integrity, Availability).

3.4 Interviewing Students

We conducted two types of student interviews: talk-aloud interviews to uncover student misconceptions [45], and cognitive interviews as part of the validation process [31]. We conducted the interviews with students from three diverse schools: UMBC (a public research university), Prince George’s Community College, and Bowie University (a Historically Black College or University (HBCU)). UMBC’s Institutional Review Board (IRB) approved the protocol.

We developed a series of scenarios based on the five core concepts identified in the Delphi processes. Before drafting complete test items, we conducted 26 one-hour talk-aloud interviews to uncover misconceptions, which we subsequently used to generate distractors. During the interviews we asked open-ended questions of various difficulties based on prepared scenarios. For each scenario,

we also prepared a “tree” of possible hints and follow-up questions, based on the student’s progress. The interviewer explained that they wanted to understand how the student thought about the problems, pointing out that the interviewer was not an expert in cybersecurity and that they were not evaluating the student. One or two cybersecurity experts listened to each interview, but reserved any possible comments or questions until the end. We video- and audio-recorded each interview.

We transcribed each interview and analyzed it using novice-led paired thematic analysis [45]. Labeling each section of each interview as either “correct” or “incorrect,” we analyzed the data for patterns of misconceptions. Four themes emerged: overgeneralizations, conflated concepts, biases, and incorrect assumptions [45]. Together, these themes reveal that students generally failed to grasp the complexity and subtlety of possible vulnerabilities, threats, risks, and mitigations.

As part of our validation studies, we engaged students in cognitive interviews during which a student reasoned aloud while they took the CCI or CCA. These interviews helped us determine if students understood the questions, if they selected the correct answer for the correct reason, and if they selected incorrect answers for reasons we had expected. These interviews had limited contributions since most subjects had difficulty providing rationales for their answer choices. The interviews did reveal that specific subjects had difficulty with some of the vocabulary, prompting us to define selected terms (e.g., masquerade) at the bottom of certain test items.

Although there is significant value in conducting these interviews, they are a lot of work, especially analysis of the talk-aloud interviews. For the purpose of generating distractors, we now recommend very strongly the simpler technique of asking students (including through crowdsourcing) open-ended stems, without providing any alternatives (see Sect. 3.7).

3.5 Creating Scenarios

To prepare for our initial set of interviews (to uncover student misconceptions), we created several interview prompts, each based on an engaging scenario. Initially we created twelve scenarios organized in three sets of four, each set including a variety of settings and difficulty levels.

We based our first CCI test items on the initial twelve scenarios, each test item comprising a scenario, stem, and five answer choices. Whenever possible, to keep the stem as simple as possible, we placed details in the scenario rather than in the stem. Initially, we had planned to create several stems for each scenario, but as we explain in Sect. 4, often this plan was hard to achieve. Over time, we created many more scenarios, often drawing from our life experiences. Sometimes we would create a scenario specifically intended to target a specific concept (e.g., identify the attacker) or topic (e.g., cyberphysical system).

For example, one of our favorite CCI scenarios is a deceptively simple one based on lost luggage. We created this scenario to explore the concept of identifying targets and attackers.

Lost Luggage. *Bob’s manager Alice is traveling abroad to give a sales presentation about an important new product. Bob receives an email with the following message: “Bob, I just arrived and the airline lost my luggage. Would you please send me the technical specifications? Thanks, Alice.”*

Student responses revealed a dramatic range of awareness and understanding of core cybersecurity concepts. Some students demonstrated lack of adversarial thinking in suggesting that Bob should simply e-mail the information to Alice, reflecting lack of awareness of potential threats, such as someone impersonating Alice or eavesdropping on the e-mail. Similarly, others recognized the need to authenticate Alice, but still recommended e-mailing the information without encryption after authenticating Alice. A few students gave detailed thoughtful answers that addressed a variety of concerns including authentication, confidentiality, integrity, policy, education, usability, and best practices.

We designed the CCA for subjects with greater technical sophistication, for which scenarios often include an artifact (e.g., program, protocol, log file, system diagram, or product specification). We based some CCA test items directly on CCI items, adding an artifact. In most cases we created entirely new scenarios. In comparison with most CCI scenarios, CCA scenarios with artifacts require students to reason about more complex real-world challenges in context with specific technical details, including ones from the artifact. For example, inspired by a network encountered by one of our team members, the CCA switchbox scenario (Fig. 1) describes a corporate network with switchbox. We present this scenario using prose and a system diagram and use it to target the concept of identifying security goals. As revealed in our cognitive interviews, these artifacts inspired and challenged students to apply concepts to complex situations. A difficulty in adding artifacts is to maintain focus on important timeless concepts and to minimize emphasizing particular time-limited facts, languages, or conventions.

Responses from our new crowdsourcing experiment (Sect. A) suggest that some subjects were confused about how many LANs could be connected through the switch simultaneously. Consequently, we made one minor clarifying edit to the last sentence of the scenario: we changed “switch that physically connects the computer to the selected LAN” to “switch that physically connects the computer to exactly one LAN at a time.”

Switchbox. *A company has two internal Local Area Networks (LANs): a core LAN connected to an email server and the Internet, and an accounting LAN connected to the corporate accounting server (which is not connected to the Internet). Each desktop computer has one network interface card. Some computers are connected to only one of the networks (e.g., Computers A and C). A computer that requires access to both LANs (e.g., Computer B) is connected to a switchbox with a toggle switch that physically connects the computer to exactly one LAN at a time.*

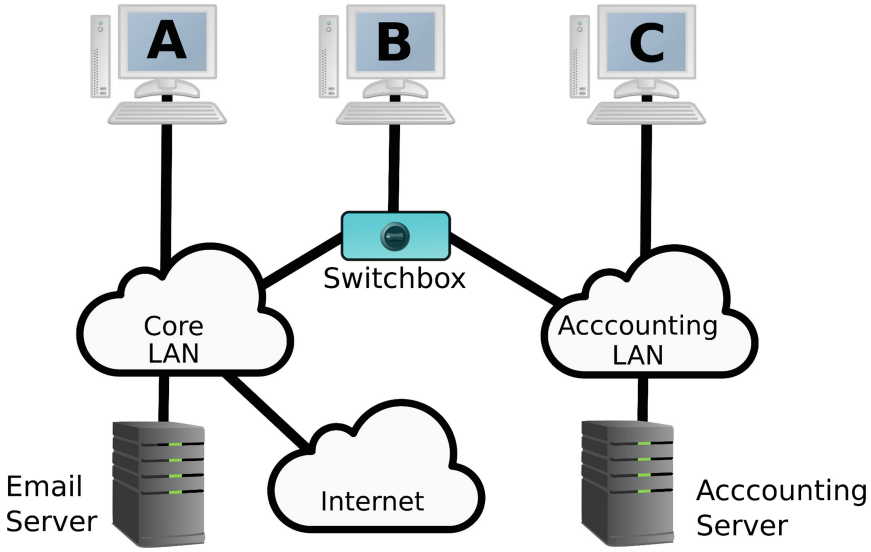


Fig. 1. CCA switchbox scenario, which includes an artifact of a network diagram with switchbox.

Comparing the CCI “lost luggage” scenario to the CCA “switchbox” scenario, one can see that the CCI scenario is simple, requiring few details to be clear. On the other hand, the CCA scenario requires the Consideration and analysis of a greater number of facts and properties of the system. Some of these facts, such as “[the accounting LAN] is not connected to the Internet” and that “each desktop computer has one network interface card,” may have been added in discussion as the problem developers required clarification in their discussion of the scenario. In conjunction with the artifact, the scenario serves to constrain the problem in such a way that the system can be well-understood.

3.6 Drafting Stems

Drafting a stem requires careful consideration of several points, in the context of the scenario and alternatives. Each test item should primarily target one of the five core concepts, though to some degree it might involve additional concepts. The stem should be meaningful by itself, and an expert should be able to answer it even without being provided any of the alternatives. We try to keep each stem as focused and short as reasonably possible. To this end, we try to place most of the details into the scenario, though stems may add a few supplemental details. Each test item should measure conceptual understanding, not informational knowledge, and not intelligence. Throughout we consider the “Vanderbilt” guidelines [2], which, among other considerations, caution against negatively worded stems, unless some significant learning outcome depends on such negativity (e.g., “What should you **NOT** use to extinguish an oil fire?”).

There are many pitfalls to avoid, including unclear wording, ambiguity, admitting multiple alternatives, using unfamiliar words, and being too easy or too hard. As a rule of thumb, to yield useful information, the difficulty of each test item should be set so that at least 10%, and at most 90%, of the subjects answer correctly. We try hard to leave nothing to the subject's imagination, making it a mistake for subjects to add details or assumptions of their own creation that are not explicitly in the scenario or stem.

To carry out the detailed work of crafting test items, we created a problem development group, whose regular initial members were cybersecurity experts Sherman, Golaszewski, and Scheponik. In fall 2018, Peterson joined the group. Often during our weekly CATS conference calls, we would present a new CCI item to Herman and Oliva, who are not cybersecurity experts. It was helpful to hear the reactions of someone reading the item for the first time and of someone who knows little about cybersecurity. An expert in MCQs, Herman was especially helpful in identifying unintentional clues in the item. Herman and Oliva were less useful in reviewing the more technical CCA test items.

We created and refined stems in a highly iterative process. Before each meeting, one member of the problem development group would prepare an idea, based strongly on one of our core concepts. During the meeting, this member would present their suggestion through a shared Google Doc, triggering a lively discussion. Other members of the group would raise objections and offer possible improvements, while simultaneously editing the shared document. Having exactly three or four members present worked extremely well for us, to provide the diverse perspectives necessary to identify and correct issues, while keeping the discussion controlled enough to avoid anarchy and to permit everyone to engage. Over time we became more efficient and skilled at crafting test items, because we could better overcome predictable difficulties and avoid common missteps.

Sometimes, especially after receiving feedback from students or experts, we would reexamine a previously drafted test item. Having a fresh look after the passage of several weeks often helped us to see new issues and improvements.

Continuing the switchbox example from Sect. 3.5, Fig. 2 gives three versions of this CCA stem during its evolution. In Version 1, we deemed the open-ended phrasing as too subjective since it is impossible for the subject to determine definitely the network design's primary intent. This type of open-ended stem risks leading to multiple acceptable alternatives, or to one obviously correct alternative and four easily rejected distractors. Neither of these outcomes would be acceptable.

In Version 2, instead of asking about the designer's intent, we ask about security goals that this design supports. We also settled on a unified language for stems, using the verb "choose," which assertively emphasizes that the subject should select one best answer from the available choices. This careful wording permits the possibility that the design might support multiple security goals, while encouraging one of the security goals to be more strongly supported than the others.

Seeking even greater clarity and less possible debate over what is the best answer, we iterated one more time. In Version 3, we move away from the possibly

Version 1: What security goal is this design primarily intended to meet?
 Version 2: Choose the security goal that this design best supports.
 Version 3: Choose the action that this design best prevents.

Fig. 2. Evolution of the stem for the CCA switchbox test item.

subjective phrase “goal that this design best supports” and instead focus on the more concrete “action that this design best prevents.” This new wording also solves another issue: because the given design is poor, we did not wish to encourage subjects to think that we were praising the design. This example illustrates the lengthy, careful, detailed deliberations we carried out to create and refine stems.

3.7 Developing Distractors

Developing effective distractors (incorrect answer choices) is one of the hardest aspects of creating test items. An expert should be able to select the best answer easily, but a student with poor conceptual understanding should find at least one of the distractors attractive. Whenever possible, we based distractors on misconceptions uncovered during our student interviews [45]. Concretely doing so was not always possible in part because the interviews did not cover all of the scenarios ultimately created, so we also based distractors on general misconceptions (e.g., encryption solves everything). The difficulty is to develop enough distractors while satisfying the many constraints and objectives.

For simplicity, we decided that each test item would have exactly one best (but not necessarily ideal) answer. To simplify statistical analysis of our assessments [31], we decided that each test item would have the same number of alternatives. To reduce the likelihood of guessing correctly, and to reduce the required number of test items, we also decided that the number of alternatives would be exactly five. There is no compelling requirement to use five; other teams might choose a different number (e.g., 2–6).

Usually, it is fairly easy to think of two or three promising distractors. The main difficulty is coming up with the fourth. For this reason, test creators might choose to present four rather than five alternatives. Using only four alternatives (versus five) increases the likelihood of a correct guess; nevertheless, using four alternatives would be fine, provided there are enough test items to yield the desired statistical confidence in student scores.

As we do when drafting stems, we consider the “Vanderbilt” guidelines [2], which include the following: All alternatives should be plausible (none should be silly), and each distractor should represent some misconception. Each alternative should be as short as reasonably possible. The alternatives should be mutually exclusive (none should overlap). The alternatives should be relatively homogeneous (none should stand out as different, for example, in structure, format, grammar, or length). If all alternatives share a common word or phrase, that phrase should be moved to the stem.

Care should be taken to avoid leaking clues, both within a test item and between different test items. In particular, avoid leaking clues with strong diction, length, or any unusual difference among alternatives. Never use the alternatives “all of the above” or “none of the above;” these alternatives complicate statistical analysis and provide little insight into student understanding of concepts. If a negative word (e.g., “**NOT**”) appears in a test item, it should be emphasized to minimize the chance of student misunderstandings. As noted in Sect. 3.6, typically stems should be worded in a positive way.

To develop distractors, we used the same interactive iterative process described in Sect. 3.6. We would begin with the correct alternative, which for our convenience only during test item development, we always listed as Alternative A. Sometimes we would develop five or six distractors, and later pick the four selected most frequently by students. To overcome issues (e.g., ambiguity, possible multiple best answers, or difficulty coming up with more distractors), we usually added more details to the scenario or stem. For example, to constrain the problem, we might clarify the assumptions or adversarial model.

Reflecting on the difficulty of conducting student interviews and brainstorming quality distractors, we investigated alternate ways to develop distractors [38, 39]. One way is to have students from the targeted population answer stems *without* being offered any alternatives. By construction, popular incorrect answers are distractors that some subjects will find attractive. This method has the advantage of using a specific actual stem. For some test items, we did so using student responses from our student interviews. We could not do so for all test items because we created some of our stems after our interviews.

An even more intriguing variation is to collect such student responses through crowdsourcing (e.g., using Amazon Mechanical Turk [27]). We were able to do so easily and inexpensively overnight [38, 39]. The main challenges are the inability to control the worker population adequately, and the high prevalence of cheaters (e.g., electronic bots deployed to collect worker fees, or human workers who do not expend a genuine effort to answer the stem). Nevertheless, even if the overwhelming majority of responses are gibberish, the process is successful if one can extract at least four attractive distractors. Regardless, the responses require grouping and refinement. Using crowdsourcing to generate distractors holds great promise and could be significantly improved with verifiable controls on the desired workers.

Continuing the switchbox example from Sects. 3.5 and 3.6, we explain how we drafted the distractors and how they evolved. Originally, when we had created the scenario, we had wanted the correct answer to be preventing data from being exfiltrated from the accounting LAN (Alternative D is a more specific instance of this idea). Because the system design does not prevent this action, we settled on the correct answer being preventing access to the accounting LAN. To make the correct answer less obvious, we worded it specifically about employees accessing the accounting LAN from home. Intentionally, we chose not to use a broader wording about people accessing the accounting LAN from the Internet, which subjects in our new crowdsourcing experiment (Sect. A) subsequently came up with and preferred when presented the open-ended stem without any alternatives.