VALIDATING A CONCEPT INVENTORY FOR CYBERSECURITY

BY

SPENCER OFFENBERGER

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Adviser:

Professor Michael Loui

ABSTRACT

The validity of a concept inventory is imperative to its adoption by the education community. In this thesis, we evaluate the validity of the Cybersecurity Concept Inventory for assessing student knowledge of core cybersecurity concepts after a first course on the topic. Our evaluation involved expert review and student performance data. A panel of 12 experts in cybersecurity reviewed each item of the Cybersecurity Concept Inventory (CCI), and the majority agreed that every item measured appropriate cybersecurity knowledge. We gave the CCI to 142 students from six different institutions taking a first cybersecurity course either online or proctored by the professor of the course. We used Classical Test Theory to evaluate the quality of the CCI. This evaluation showed that the CCI is sufficiently reliable for measuring students' knowledge of cybersecurity and that the CCI may be too difficult as a whole. We describe the results of both the expert review and the pilot test in further detail and provide recommendations for the continued improvement of the CCI.

To my parents, for their love and support.

ACKNOWLEDGMENTS

Thank you to the CATS team including Alan Sherman at University of Maryland Baltimore County (UMBC), Peter Peterson at University of Minnesota Duluth (UMD), and Linda Oliva at UMBC. Thank you to Dr. Herman for the guidance and help throughout this thesis. Thank you to Dr. Loui for taking the time to help and for making me a better writer. Thank you to the experts and professors who reviewed and administered the CCI: Dr. Arno Wacker at UMBC, Dr. Yu Cai at Michigan Technological University (MTU), Dr. Filipo Sarevski at DePaul University, Joe Roundy at Montgomery College, Dr. Jelena Mirkovic at USC, Dr. Ankur Chattopadhyay at University of Wisconsin Green Bay, Simson Garfinkel at ACM, Dr. Tamara Denning at Utah, Dr. William Butler at Capitol Tech, Dr. Denning Grauvogel at Utah, Philip Ritchey at Texas A&M, and Nicole Hands at Purdue Polytechnic.

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 BACKGROUND 3

CHAPTER 3 METHODS 9

CHAPTER 4 RESULTS 12

CHAPTER 5 DISCUSSION 20

CHAPTER 6 CONCLUSION 25

REFERENCES 26

APPENDIX A ITEM DETAILS 29

APPENDIX B FULL ASSESSMENT 30

LIST OF TABLES

2.1 Five Core Concepts of Cybersecurity 4

3.1 Breakdown of Students by University 11

4.1 Comparison with Other Instruments 14
4.2 Instrument Statistics 14
4.3 Cronbach's α Analysis 15
4.4 Difficulty and Discrimination of Each Item 16
4.5 Cronbach's α by Concept 16
4.6 Distractor Distribution 18
4.7 Distractor Discrimination Values 19

A.1 Information About CCI Items 29

LIST OF FIGURES

2.1 Validity of Discrimination and Difficulty 7

4.1 Expert Response to Items 13
4.2 Difficulty vs. Discrimination 17
4.3 Concept Alignment of Top Quartile 17

CHAPTER 1

INTRODUCTION

1.1 Motivation

Computers are becoming ubiquitous: they are used in diverse contexts including medical equipment, cars, and appliances. Due to this ubiquity, both cybersecurity experts and non-experts involved in these fields need to understand the core concepts of cybersecurity. For example, after a string of ransomware attacks targeted hospitals, Wirth called for healthcare technology management professionals to undertake cybersecurity training [1]. The number of fields requiring basic cybersecurity concepts will likely continue to rise as attackers' targets expand.

Because the industries being affected by cyberattacks are expanding, security professionals are in high demand. This demand cannot be met with the current levels of cybersecurity education, however. Libicki et al. [2] go as far as to say "the shortage of cybersecurity experts in the federal government is serious to the point of being a national security threat to the United States." Despite the importance and ubiquity of cybersecurity, there is little research on how to teach cybersecurity effectively. A valid and broadly used conceptual instrument for cybersecurity is a vital resource for supporting rigorous research on the efficacy of various teaching methods for cybersecurity education. Unfortunately, at this time, there are no validated research instruments to assess students' conceptual knowledge of cybersecurity.

Sherman et al. began the Cybersecurity Assessment Tools (CATS) project to meet this need for validated research instruments to assess the effectiveness of cybersecurity education [3–8]. The CATS Project is developing two concept inventories (CIs) to evaluate how well teaching practices help students learn core cybersecurity concepts: the Cybersecurity Concept Inventory (CCI) and Cybersecurity Curriculum Assessment (CCA). The CCI as-

sesses how well a student has learned the basic concepts of cybersecurity after one cybersecurity course. The CCA assesses how well a student has learned cybersecurity concepts after completing a full cybersecurity curriculum.

1.2 Validity and Concept Inventories

CIs have been used to show that students regrettably succeed in traditional assessments through fact memorization rather than conceptual understanding [9–11]. With a deeper conceptual understanding, students can learn more efficiently in the future and can transfer their knowledge across contexts [11]. CIs have been effectively used to promote the adoption of evidence-based teaching practices across STEM that are conducive to students developing a deeper conceptual understanding [9, 10, 12].

A CI can be powerful and useful only if it is deemed as a valid instrument by the education community that will use the instrument. A CI is valid if it effectively evaluates targeted concepts and can be used to draw a reasonable inference of student knowledge [13]. The validity of the instrument is established by a set of evidence and arguments about whether the instrument can be appropriately used to draw these inferences. To establish the validity of our instrument, we are following the design and evaluation framework recommended by the National Research Council [14, 15].

1.3 Outline of Thesis

In this thesis, we review the development process of the CCI, and we compare that process with the development of other CIs. We then describe the framework we use to evaluate whether the CCI can be used validly to assess student knowledge of cybersecurity concepts. We then describe the research methods for the expert panel review and pilot test with students. We analyze the results of this pilot test using Classical Test Theory (CTT). We then discuss these findings to identify the strengths of the CCI and to recommend future improvements for the CCI.

CHAPTER 2

BACKGROUND

The National Research Council recommends establishing a cognitive framework for the design of an instrument [15]. This cognitive framework defines what knowledge of a concept should be assessed and the ways in which students reveal their knowledge, or lack of knowledge, about that topic. Prior work on the CATS Project has focused on establishing this cognitive framework, providing baseline arguments for the validity of the CCI.

Because a test cannot be universally valid for every population or purpose, we need to define carefully the contexts, populations, and purposes for which the CCI is valid. We intended the CCI to measure the cybersecurity conceptual knowledge of students who have completed a first course in cybersecurity. Cybersecurity is taught to an increasingly wide range of stakeholders, including policymakers, computer scientists, medical professionals, and business professionals. Each stakeholder's courses vary in focus and depth. Because of this high variance, we have chosen to optimize the CCI for the largest population of cybersecurity professionals—computer scientists. While the CCI may provide useful insights about the conceptual knowledge of policymakers or others, our goal is to have the instrument provide the most insight about computer science students.

2.1 CATS Project

In accordance with the recommendations of the National Research Council, we based the design of the CCI on the consensus opinions of a panel of experts and on documented student misconceptions [3, 4, 6].

Parekh et al. [3] began the CATS Project development by identifying the core concepts of cybersecurity using a Delphi process. A Delphi process is a rigorous and structured method for creating consensus among experts about

potentially contentious issues, such as what subset of concepts should be included on the CCI [16]. A Delphi process has been used to identify the cognitive framework of several previous CIs [17]. The Delphi process for the CCI identified five concepts all related to adversarial thinking to include in the CCI seen in Table 2.1 [3]. From these concepts, Sherman et al. [5] developed cybersecurity scenarios that require students to understand these concepts. For example, a scenario that covers the concept Confidentiality, Integrity, Availability, and Authentication Attacks (C) involves a hypothetical government facility where we define defenses and biometric authentication methods. This scenario defines the defenses that allow for questions on potential attacks that could exploit the defenses.

Table 2.1: Five Core Concepts of Cybersecurity

Identify Vulnerabilities and Failures (V)
Confidentiality, Integrity, Availability, and Authentication Attacks (C)
Devise a Defense (D)
Identify the Security Goals (G)
Identify Potential Targets and Attackers (T)

Using these scenarios, Scheponik et al. [4] performed think-aloud interviews to discover students' misconceptions and problematic reasoning about cybersecurity [6]. Example forms of problematic reasoning include students' beliefs that encryption protects against most any cybersecurity threat and the belief that cybersecurity threats come only from outside an organization.

Using findings from these interviews and the scenarios developed by Sherman et al. [5], we created the CCI multiple-choice questions, called *items*. Each CCI item consists of a scenario, a stem (i.e., a question about the scenario), and five answer options. One option is correct, and the other four are incorrect. The incorrect options are called *distractors* and are based on the interview findings. A student answers an item by choosing one option. If the student chooses the correct option, the student's answer is correct. If the student chooses a distractor or does not choose an option, the answer is considered incorrect. The student's score on the CCI is the total number of items that the student answers correctly. This score ranges from 0 to 25.

In this thesis, we followed the National Research Council's recommended development process. We assembled a panel of 12 experts to review whether the draft CCI indeed matched the targeted cognitive framework. Once an instrument is created, it should be administered to its targeted demographic

and be statistically evaluated [15]. We administered a pilot test of the CCI to a group of 142 students from six universities to evaluate whether students responded to items on the CCI according to our expectations from the interviews. We use statistical analysis of student responses to determine what inferences can be validly drawn from administrations of the CCI.

2.2 Classical Test Theory (CTT)

Jorion et al. [18] outlined three basic properties of a valid CI: CI indicates overall understanding of the concepts, CI indicates understanding of a specific concept, CI indicates misconceptions or student errors. Jorion et al. recommended using a series of statistical tests to demonstrate whether a CI possesses these properties [18]. CTT is often the first evaluation paradigm used to evaluate an instrument because it is useful with small sample sizes [19]. CTT is more practical than more exhaustive analytics such as Item Response Theory (IRT) because CTT allows us to find problematic items and distractors and suggest modifications with fewer students than required for an IRT analysis. This analysis enables rapid iteration and improvement of the CI.

According to CTT, an assessment instrument should minimize error. All of the instrument's items should test a single construct. Each item should be neither too hard nor too easy. Each item should provide a good estimate of the student's overall ability

2.3 Reliability

Reliability is a measure of the likelihood that repeated measurements of the same student will yield the same score. If an instrument is not reliable, it cannot be valid.

In CTT, the core assumption is that a student's *observed score (X)* consists of two hypothetical values: a student's *true score (T)* and some random *error (E)* [19]. This model is expressed symbolically as $X = T + E$ [17]. If a test is not biased, then the average error is zero. As a consequence, the student's true score would be the average of an infinite number of administrations of

the text [20]. A reliable instrument minimizes the error so the observed score best reflects the student's understanding.

The conventional measurement used for internal reliability is Cronbach's α. Cronbach's α is "an estimate of the correlation between two random samples of items from a universe of items like those in the test" [21]. We can determine Cronbach's α without administering the CCI multiple times if two conditions are met: (1) the instrument measures a single construct, and (2) the item is either correct or incorrect [17]. A reliable instrument will lead to α values that are close to 1.

There is no universally acceptable Cronbach value, but 0.8 is considered good and 0.7 is the minimum value considered satisfactory according to Panayides [22] and Jorion et al. [18].

Cronbach's α is also used to coarsely evaluate the quality of each item. The addition of each item should increase the overall reliability of the instrument [17]. By removing an item and then recalculating the α value, we can judge the quality of that specific item. When the removal of an item increases the value of α, the item is particularly poor and should be removed.

The standard error is a function of α and defines a confidence interval for each student's true score. We calculate standard error using

$$SE = S_x\sqrt{1-\alpha}$$

where S_x is the standard deviation of the sample and α is the Cronbach's α. When the standard error is small, we can be confident that students with different observed scores have different true scores.

2.4 Difficulty and Discrimination

Reliability alone does not indicate the instrument provides a valid representation of student knowledge. The validity of the instrument can be further established by examining each item's difficulty and discrimination. The difficulty of an item is the fraction of students with the correct response [19]. Each item of the instrument should have a balanced range of difficulties falling within 0.2 to 0.8 [17, 18]. When the difficulty is outside this range, it does not effectively separate students with different levels of understanding.

The discrimination of an item is the point-biserial correlation between the item and the overall performance [18]. When an item's discrimination is low, weaker students (low total scores) perform similarly to stronger students (high total scores) on that item. A good item will have a discrimination of at least 0.2 [17]. An item with proper discrimination and difficulty will fall within the largest square area as noted by the dotted lines in Figure 2.1.

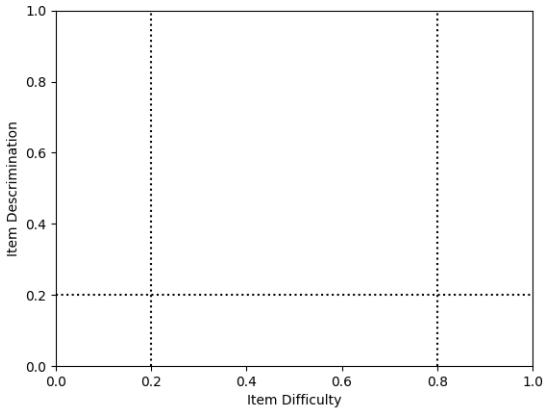


Figure 2.1: Validity of Discrimination and Difficulty

2.5 Distractor Analysis

Distractor analysis utilizes the information about the answers choices students chose to better identify why including an item did not improve α or why the item has a difficulty and discrimination outside the accepted range. To analyze distractors we split the students into tertiles (thirds) according to total scores. After splitting the students, we take the proportion of students selecting each response [19]. There are certain trends we expect to see: (1) the percentage of students selecting the correct answer should increase from the bottom third to the top third, (2) the item's difficulty for the top third of students should be near 0.8, (3) each distractor should have a negative discrimination value [23]. The discrimination value of a distractor is the discrimination when the distractor is considered to be the correct answer. The discrimination value of the correct answer is the same as the discrimination of the item.

2.6 Concept Subtests

Cronbach's α can be applied to a group of items called a *subtest*. In our case, we propose that there may be 5 subtests in the CCI, each aligning with the five concepts identified in the Delphi process. We evaluate these subtests separately to assess reliability to determine whether we can interpret understanding of the concepts from these subtests alone. Ideally, each subtest should have a reliability similar to the overall instrument. In practice, having a similar reliability to the entire instrument is difficult because each subtest has fewer items and the value of Cronbach's alpha tends to increase with the number of items.

CHAPTER 3

METHODS

3.1 Expert Panel

The *initial CCI* comprised the 32 items developed using the processes described earlier. We gave these items to an expert panel consisting of 11 professors with backgrounds in cybersecurity and one cybersecurity professional for review. The experts each received the initial CCI online containing each of the items. We asked experts to rate each item on the scale Accept, Accept with Minor Revisions, Accept with Major Revisions, and Reject, and to comment on the item. After they answered the item, experts were shown the correct answer and given the option to provide additional comments on the correct answer.

The expert comments mainly concerned the clarity of items. We addressed their comments and then checked back with them for approval of the updated items. An example of how we addressed these comments is in Section 4.1.

For some items, the experts disagreed with the content or the correct answer. When the experts disagreed, we omitted that item from the CCI. After we removed these items, the experts' reviews were used to rate the remaining items. The highest rated items were incorporated into the current CCI.

During the expert reviews, we continued developing items for the CCI. These items were intended to be used as alternates. One alternate item, Q25, was not reviewed by experts but was included. We thought this item evaluated the core concept C better than those that were reviewed. This item and those with the best reviews form the current 25 item CCI.

3.2 Current CCI

We selected items with a range of difficulties based on our best estimation. There were six easy items, 16 medium items, and three hard items. The actual performance of students would likely differ from our estimations. Each item covers one of the five major concepts shown in Table 2.1. The items are shown in Table A.1 in Appendix A. This table shows the name, concept, topic of each item in the current CCI.

3.3 Pilot Test

The goal of the pilot test was to administer the current CCI to a small group of 100-200 students and then use the results of this pilot test to suggest modifications to the instrument. We concluded the pilot test in December 2018 after 142 students from six universities completed the CCI.

Professors at each university had the option of administering a paper version or online version of the CCI. Both versions included the instructions seen in Appendix A. The distractors, scenarios, and questions were identical in both versions.

A professor proctored the paper version of the CCI by allocating 50 minutes for students to take the 25 item CCI in class. Students then completed the CCI to the best of their abilities. The professor collected the CCI papers and sent them to us. Then we recorded each student's answer to every item.

If the professor decided to administer the online version, students were provided a link to the CCI. The online version differed from the paper version in three ways. First, the online version had a random ordering of distractors. Second, items that shared a scenario were randomly ordered within that scenario. For example, if Q1 and Q2 are the two items in the one scenario, Q1 could appear before or after Q2 but always together with it. The reason for randomizing the online version was to dissuade collusion between students and to minimize any possible effect of item order on student performance. Because students who had access were all in the same course they may have attempted to work together even if they received no benefit from receiving a better score. Third, there was no hard time limit. Students were told to spend 50 minutes but this was not strictly enforced. The student completed

the online version and then selected a submit button to save and submit their CCI.

3.4 Pilot Demographics

The universities included in the pilot test have diverse locations and populations. Universities A and D are large Midwestern public universities. Each has over 40 thousand students enrolled. University E is a large public university from the Southwest with over 40 thousand students enrolled. Universities B, C, F are smaller universities from the Midwestern and Eastern part of the United States. These Universities have 10 thousand or fewer students enrolled.

The demographics of the study including institution and response rate are in Table 3.1. University A was the only group given the CCI in paper format. At University D, a link to the instrument was sent to six members of a professional engineering club who were taking the course. At the other universities, a professor sent a link to the instrument to the students in the course.

Table 3.1: Breakdown of Students by University

University	Number of Students	Potential Number of Students	Response Rate (%)
University A	91	120	76
University B	12	20	60
University C	1	12	16
University D	6	6	100
University E	17	50	34
University F	12	20	60
No University Specified	3		
Total	142	228	62

3.5 Incomplete Submissions

Two types of null responses were discarded. The first was 12 responses in which no effort was made to complete the CCI, leaving all items blank. The second was three response from a student who began the CCI but did not continue and terminated it early (completing fewer than 5 items). If a student completed the majority of the items but left a minority blank, their results were included.

CHAPTER 4

RESULTS

In this chapter, we present results from the expert review of the CCI and our psychometric analysis of students' responses to the CCI. To help the reader interpret our findings, we compare our results with three CIs evaluated with the same techniques. These CIs are the Concept Assessment Tool for Statics (27 questions and 1,372 students), the Statistics Concept Inventory (38 questions and 402 students), and the Dynamics Concept Inventory (29 questions and 5,966 students) [18]. We chose these CIs because they are the few technical CIs that analyzed the CIs using similar techniques.

4.1 Expert Panel

The expert panel consisted of 12 experts in cybersecurity or a related field. These experts reviewed each of the individual items and rated them on a scale of Accept, Accept with Minor Revisions, Accept with Major Revisions, and Reject. The results of this review process can be seen in Figure 4.1. Although 32 items were reviewed, the figure presents the results for the 24 items in the current CCI. The items selected for the CCI were reviewed positively receiving a vast majority of Accept and Accept with Minor Revisions.

Additionally, experts left comments for each item which we used to revise the items that received Reject ratings. As an example, we show how we used expert reviews to revise item Q4. Q4 covered a potential SQL injection vulnerability and the means of defending against it. The initial wording of the Q4 scenario is below.

> **Scenario A3** When a user Mike O'Brien registered a new account for an online shopping site, he was required to provide his username, address, first and last name, and a password. Immediately after Mike submitted his request, you—as the security

engineer—receive a database input error message in the logs.

Experts commented that this wording is imprecise because an error in the logs is not "received" but rather written into the log on the server. The word "received" implies the error was noticeable and could lead students to infer that the error came from the client-side. The item was modified to replace "receive a database input error" with "observe a database input error." The change makes it clear that the user input did not cause an alert and that the message was logged on the server-side. The clarification will lead students away from client-side solutions such as "more thoroughly test the software before deploying it" and toward server-side solutions such as the correct response, "sanitize input at the server side." The expert review process strengthened the clarity which is critical to measuring a student's conceptual knowledge.

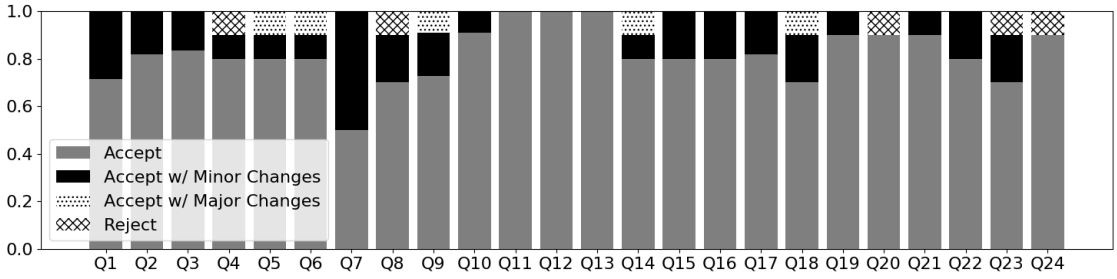


Figure 4.1: Expert Response to Items

4.2 Reliability and Standard Error

Cronbach's α is a measure of the reliability of the instrument. The Cronbach's α of the CCI in this pilot test is 0.78 as seen in Table 4.2. This α is close to Jorion et al.'s recommendation for good reliability of 0.80 and above Panayiotis's minimum recommendation of 0.70. The reliability of the CCI is similar to other CIs as seen in Table 4.1. The reliability of the CCI suggests that it is sufficiently reliable to be a valid CI.

The standard error of measurement defines a confidence interval for each student's true score. The standard measurement error of the CCI was 2.13 for this pilot test. A 2.13 standard error implies the 68% confidence interval

for a student's true score, given that a mean observed score of 8.61 points is from 6.48 to 10.74.

Table 4.1: Comparison with Other Instruments

Measurement	**CCI**	**Statics**	**Statistics**	**Dynamics**
Cronbach's α	0.78	0.84	0.64	0.74
Minimum Difficulty Value	0.10	0.16	0.03	0.06
Maximum Difficulty Value	0.66	0.78	0.87	0.91
Minimum Discrimination Value	0.16	0.18	-0.13	0.01
Maximum Discrimination Value	0.47	0.65	-0.57	0.56
Number of Items w/ Difficulty Below 0.2	5	1	3	4

Table 4.2: Instrument Statistics

Cronbach's α	0.78
Standard Error of Measurement	2.13
Mean (Out of 25)	8.61
Standard Deviation	4.58

Cronbach's α can be used as a coarse evaluation of the quality of an item. The addition of each item should increase the quality of the instrument, and excluding that item should decrease the overall reliability. Table 4.3 shows the results of the Cronbach calculation with each item excluded. Ideally, for each item the value of α with that item deleted should be less than the overall value of α, 0.78. There are no items that decrease the overall reliability and consequently need to be removed.

4.3 Difficulty and Discrimination

The difficulty of an item is the fraction of students with the correct response. If an item is too hard, the item is separating only strong students from strong students. If an item is too easy, it cannot differentiate any students. The acceptable range of difficulty is between 0.20 and 0.80. The difficulty of each item can be seen in Figure 4.2 and Table 4.4. The range of difficulty for the CCI is 0.10 to 0.66. When compared to the other instruments the CCI seen in Table 4.1, it is more difficult and will have less discriminatory power. The CCI instrument overall is too difficult as evidenced by 21 out of 25 items having difficulty below 0.50 and 5 items falling outside the minimum

Table 4.3: Cronbach's α Analysis

Item Excluded	Cronbach's α	Item Excluded	Cronbach's α
Q1	0.77	*Q14*	0.76
Q2	0.76	*Q15*	0.77
Q3	0.78	*Q16*	0.76
Q4	0.76	*Q17*	0.76
Q5	0.76	*Q18*	0.77
Q6	0.77	*Q19*	0.77
Q7	0.77	*Q20*	0.77
Q8	0.77	*Q21*	0.77
Q9	0.76	*Q22*	0.76
Q10	0.77	*Q23*	0.76
Q11	0.76	*Q24*	0.77
Q12	0.76	*Q25*	0.77
Q13	0.77		

acceptable difficulty. This is more values below minimum than the other CIs seen in Table 4.1.

The discrimination of an item indicates the amount of information an item gives about the overall performance of the student. High discrimination indicates that a student's performance on a given item is highly correlated to overall performance. The acceptable range of discrimination is anything above 0.20. Figure 4.2 shows the discrimination of each item. The range of discrimination is 0.16 to 0.47. The discrimination range is not as high as that of other CIs seen in Table 4.1, but the bottom of the range is much higher than the Statistics Concept Inventory and Dynamics Concept Inventory. The other CIs had one, ten, and five items fall below the 0.20, while the CCI had three items below 0.20. The fact that most of the items are above the minimum value is an encouraging indicator that the assessment is valid.

4.4 Concept Subtests

The CCI consists of five concepts: V, C, D, G, and T. Each item assesses one of these concepts. The individual items within a concept can be grouped and the Cronbach's α calculated to evaluate the reliability of that concept subtest. The α's of the concept subtests are seen in Table 4.5. When evaluating the concepts, it is notable that all of the values are significantly less

Table 4.4: Difficulty and Discrimination of Each Item

Item	Discrimination	Difficulty	Item	Discrimination	Difficulty
Q1	0.22	0.24	*Q14*	0.32	0.25
Q2	0.32	0.33	*Q15*	0.25	0.10
Q3	0.16	0.26	*Q16*	0.35	0.59
Q4	0.46	0.52	*Q17*	0.35	0.52
Q5	0.35	0.18	*Q18*	0.19	0.31
Q6	0.23	0.22	*Q19*	0.27	0.28
Q7	0.30	0.66	*Q20*	0.22	0.14
Q8	0.21	0.19	*Q21*	0.23	0.44
Q9	0.33	0.61	*Q22*	0.47	0.34
Q10	0.19	0.40	*Q23*	0.38	0.49
Q11	0.34	0.36	*Q24*	0.30	0.40
Q12	0.36	0.24	*Q25*	0.24	0.14
Q13	0.21	0.28			

than 0.70 which is considered the minimum acceptable value for a reliable instrument [22]. These findings suggest that the subtests should not be used as evaluations of students' knowledge of the specific concepts.

Table 4.5: Cronbach's α by Concept

Concept	Cronbach's α	Items Included
V	0.22	Q1, Q3, Q11, Q17, Q21
C	0.45	Q2, Q5, Q14, Q18, Q24
D	0.47	Q4, Q6, Q13, Q19, Q23
G	0.36	Q8, Q9, Q10, Q22, Q25
T	0.50	Q7, Q12, Q15, Q16, Q20

The correlation between items in the subtests can be expressed with a correlation matrix. The correlation matrix is the correlation coefficient of each item with every other item. A heat map of the correlation matrix can be seen in Figure 4.3. The square regions enclose the items in a specific concept subtest. Items within the same concept subtest should be expected to correlate strongly with each other compared to items outside their subtest. We do not see these types of stronger and weaker correlations.

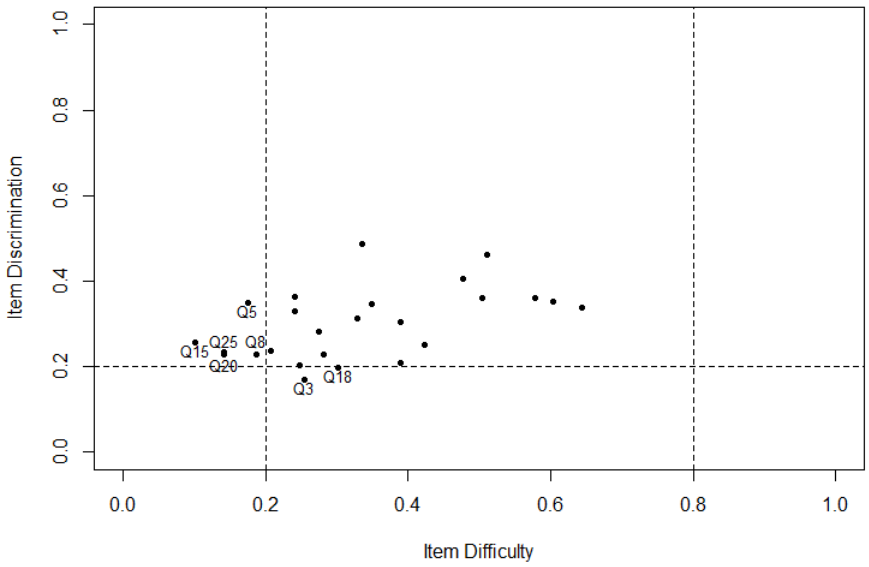
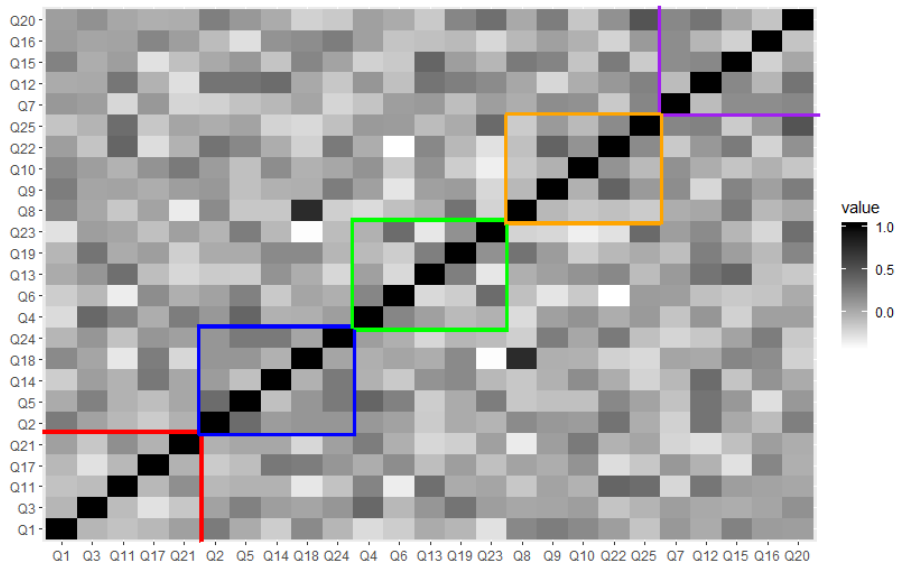


Figure 4.2: Difficulty vs. Discrimination



V: Red, C: Blue, D: Green, G: Orange, T: Purple

Figure 4.3: Concept Alignment of Top Quartile

4.5 Deeper Analysis of Specific Items

The psychometric analysis of the CCI revealed that the instrument has too many difficult items. To inform future revisions of the CCI, we are analyzing the distractor distribution and distractor discrimination to understand why some items are so difficult. We present an example of this for one of these items, Q15, which had a low difficulty of 0.10 and relatively low discrimination of 0.25 in the pilot test. We compare Q15 to a stronger item, Q4, which had a good difficulty of 0.52 and discrimination of 0.46 in the pilot test.

The distractor analysis shows the proportion of the students' answers in each tertile. The distractor analysis for both Q15 and Q4 can be seen in Table 4.6. In Q4, which has a good distribution, the percentage of students who chose the correct option increases from the bottom tertile to the top, and the top tertile generally answers the question correctly. For Q15, although the percentage of students who selected the correct option increases from the bottom tertile to the top, there is little difference between the top and middle tertiles. Additionally, the top-tertile students answered Q15 correctly 18% of the time and instead selected distractor A 59% of the time. The preference for option A among the top tertile is causing the item to be too difficult.

Table 4.6: Distractor Distribution (Asterisk Identifies Correct Option)

Q4			
Option	*Lower*	*Middle*	*Upper*
A	0.02	0	0.03
*B	0.28	0.62	0.85
C	0.11	0	0.26
D	0.37	0.14	0
E	0.22	0.24	0.10
blank	0.02	0	0

Q15			
Option	*Lower*	*Middle*	*Upper*
A	0.22	0.36	0.59
B	0.39	0.26	0.08
C	0.15	0.17	0.15
D	0.22	0.05	0
*E	0	0.17	0.18
blank	0.02	0	0

Table 4.7 shows the distractor discrimination value for both Q4 and Q15. We expect the distractors to have negative discrimination values. Q4 has negative or zero distractor discrimination values for each distractor as well as a large positive discrimination value for the correct option. Q15 does have a large positive discrimination value for the correct option and is even above the minimum acceptable value, but distractor A has a larger, positive discrimination value. This analysis reveals that the correct option is not compelling to the strongest students. The poor performance of the strongest

students suggests that the wording or structure of the answer options may be to blame. We explore this assertion more in future work.

Table 4.7: Distractor Discrimination Values (Asterisk Identifies Correct Option)

Q4	
Option	*Discrimination*
A	0
*B	0.46
C	-0.14
D	-0.26
E	-0.04

Q15	
Option	*Discrimination*
A	0.35
B	-0.19
C	0
D	-0.26
*E	0.24

CHAPTER 5

DISCUSSION

This validation study revealed the CCI instrument could be used to evaluate cybersecurity but would benefit from minor modifications. The CCI has many good properties: high reliability and strong expert consensus on the suitability of all items. Unfortunately, our findings revealed a few weaknesses of the CCI as currently constructed: low reliability for individual concepts, items that are too difficult, and too many difficult items on the instrument.

From the results of the pilot test, the CCI appears to have satisfactory reliability, especially when compared to other CIs. The Cronbach's α was 0.78, which is considered good for a CI. In addition to the CCI reliability, no item deletions decrease the overall α. This result indicates that the individual items are all measuring the same construct of cybersecurity conceptual knowledge [17]. The reliability of the instrument is necessary for the instrument to be valid but not sufficient.

Experts positively reviewed each item and provided feedback to improve the items. Experts also provided suggestions for improving the wording and distractors of each item. We used this feedback to select the 25 items that had the strongest consensus of quality from the experts. The expert reviews provide evidence for the content validity of the CCI by demonstrating that multiple cybersecurity instructors believe that the CCI items represent conceptual knowledge that students should have after a first course in cybersecurity.

The strengths of the CCI indicate that the collection of items and individual items are well designed from an instructor perspective and reliable from a student performance perspective. However, the student response data reveals that there is still room for improvement. Notably, while we designed the CCI to assess five concepts, the student performance data did not align well with these five concepts. For example, there is no consistent correlation of the items within each concept subtest. Additionally, the items that consti-

tute a subtest have low reliability; each α for the individual concept subtest is below 0.50 [18]. Because of the low reliability of the concept subtests, we cannot recommend using the concept subtests to assess students' knowledge of each concept individually.

There are two possible interpretations for this lack of correlation and reliability within the concept subtests. First, it is possible that the items were poorly designed and do not reflect the core concepts. Second, it is possible that the concepts themselves are poorly bounded, interconnected, or too complex. Given that the expert reviewers did not express any concerns about the content of the items, we argue that the second interpretation is more likely.

Our finding of low correlation among concept subtests is a common finding among previously published CIs [18]. The commonality of this finding suggests that it is generally difficult for designers of an instrument to design effective concept subtests. While most items may primarily engage students in one concept, the concepts are likely interconnected. Students need to use multiple concepts to answer each item correctly. We believe that this fact may be especially true in cybersecurity, which requires individuals to consider the motivations or capabilities of attackers, constraints or goals of defenders, and the technologies or techniques needed to mitigate risk.

Additionally, the concepts discovered in the Delphi process may be too complex and are really culminations of similar, but separate, concepts [3]. For example, concept Confidentiality, Integrity, Availability, and Authentication Attacks (C) involves four unique forms of attack. A confidentiality attack could cover attacking a secure message protocol. An availability attack could cover a denial of service attack. Both of these examples are forms of attack and both of them are very relevant to cybersecurity. A student, however, may understand mechanisms that enable secure communications and still have very little idea about denial-of-service attacks. Since each item of the CCI must be multifaceted, creating subtests will be difficult, if not intractable, without creating isomorphic, redundant questions.

If we want to create reliable and valid concept subtests, we may need to consider other models for creating them. For example, we could try narrowing the scope of concept C to just one attribute (e.g., confidentiality). This option may not be desirable because it ignores the complexity of an attacker's varied motivations. Alternatively, we could create multiple instruments that more

fully explore each of the five core concepts, but this option would dramatically increase the work and cost of creating instruments for cybersecurity. As currently constructed, the CCI provides a reliable instrument for measuring a student's overall understanding of cybersecurity, which is a much-needed first step. Future work can explore which types of future development are needed for creating these subtests.

Unlike the alignment of the concepts, a good range of difficulty is often achieved in published CIs and necessary for the instrument to be valid. The CCI is skewed to be too difficult: five items are more difficult than the recommended level of difficulty, and for 21 out of 25 items, fewer than 50% of students answered the item correctly. This degree of difficulty suggests that some items need to be made easier to improve our ability to distinguish between students with varying abilities and knowledge. Future work on the CCI must explore how to effectively make some items easier to improve the quality of the CCI.

5.1 Limitations

There are a number of limitations in the pilot test. The most notable limitation is the depth of analysis performed on the pilot test results. IRT is not practical with the number of students who took the pilot test but would enable deeper analysis. Additionally, because the Cronbach's α for each concept subtest was so low, we did not perform analyses such as Confirmatory Factor Analysis (CFA) and Exploratory Factor Analysis (EFA). These limitations are acceptable because this study is a pilot test.

There were also limitations in the number of students from each university. Ideally, there would be a similar number of representatives from the different types of universities so that the results would not be dominated by University A. The localization may have biased the findings to one university.

5.2 Future Work

We will take Q15 as a specific example of the type of modification we will make to the difficult items. Fewer than 10% of students answered Q15 cor-

rectly, far below what is acceptable for a CI.

The item covers finding vulnerabilities in a defense and falls under concept Identify Vulnerabilities and Failures (V). The scenario describes a hypothetical nuclear treaty between two countries that requires a method of securely transmitting a message from a monitoring device. Neither country trusts the other, and the design must be fair to each country. Both parties want assurances that the message is not modified. Country A wants to ensure that the message originates from the device. Country B wants to monitor the message data in real time. The premise is: "The sender applies a keyed cryptographic hash function to each message using a key distributed only to the sender, Country A, and Country B." Students are expected to find potential vulnerabilities in the suggested outputs of the device.

The options the students had for Q15 are below.

(a) The message together with a hash of the following: message and current time.

(b) The key together with a hash of the message.

(c) The message together with a hash of the message.

(d) A hash of the message.

(e) This design cannot satisfy the system requirements

Our distractor analysis revealed that the best students chose Option A more than the correct answer. This finding reveals that, as students' knowledge increased, this wrong answer became more compelling. When constructed well, each item should lead students to pick the correct answer more often as their knowledge increases.

The preference for Option A is understandable given that it is more reasonable than the other three options. Options B and D do not even send the original message so the message cannot be verified. Option A and Option C do not guarantee that the source is sending the message: since each party has the key, it can modify the message and attach a new hash. Because A has the same structure as C with the addition of time being sent, it appears to be strictly superior to C, and thus is the best option. Students must see the problems with each option and select Option E which serves as a "none of the above." Including a "none of the above" in general makes assessments

harder [24], especially because Options A and C satisfy some of the desired properties.

The problem with the item, and further "none of the above" in general, is that Option E makes no assertion. This fact leads students to pick the most reasonable of the other choices. We have modified this item, changing Option E to make an assertion. The new Option E is "The design does not work because Countries A and B can modify the message." This new wording provides a definite assertion, which students can check and conclude that the other options do not satisfy the requirements. We anticipate that this change, while minor, will make the item easier and differentiate more students.

After making similar modifications to other items, our next work is to administer the instrument to more students and reanalyze the results. With the easier items, the difficulty will cover a better range and better separate students. The range of difficulties and modification of items that are too difficult should increase the discriminatory power of the CCI and improve the CCI's validity and usefulness.

CHAPTER 6

CONCLUSION

The purpose of the expert review and pilot test was to evaluate the validity of the CCI. The expert review and pilot testing of the CCI revealed the CCI reliably tests students' knowledge of cybersecurity. At this point, the CCI could be used as an evaluation instrument, but the scores would be low. The low scores reduce the discriminatory power of the assessment. By making the CCI easier, we will be able to create an assessment that should be broadly applicable and provide useful measurements of a broad range of cybersecurity students. Further research will cover the modifications of the items and testing with more students.

REFERENCES

[1] A. Wirth, "The importance of cybersecurity training for HTM professionals," *Biomedical Instrumentation & Technology*, vol. 50, no. 5, pp. 381–383, 2016.

[2] M. Libicki, D. Senty, and J. Pollak, *Hackers Wanted: An Examination of the Cybersecurity Labor Market*. The RAND Corporation, Jan. 2014.

[3] G. Parekh, D. DeLatte, G. Herman, L. Oliva, D. Phatak, T. Scheponik, and A. T. Sherman, "Identifying core concepts of cybersecurity: Results of two delphi processes," *IEEE Transactions on Education*, vol. 61, pp. 11–20, Feb. 2018.

[4] T. Scheponik, A. T. Sherman, D. DeLatte, D. Phatak, L. Oliva, J. Thompson, and G. L. Herman, "How students reason about cybersecurity concepts." *IEEE Frontiers in Education Conference (FIE)*, pp. 1–5, Oct. 2016.

[5] A. T. Sherman, D. DeLatte, M. Neary, L. Oliva, D. Phatak, T. Scheponik, G. L. Herman, and J. Thompson, "Cybersecurity: Exploring core concepts through six scenarios," *Cryptologia*, vol. 42, no. 4, pp. 1558–1586, Sept. 2018.

[6] J. Thompson, G. Herman, T. Scheponik, L. Oliva, A. T. Sherman, and E. Golaszewski, "Student misconceptions about cybersecurity concepts: Analysis of think-aloud interviews," *Journal of Cybersecurity Education, Research and Practice*, vol. 2018, no. 1, pp. 1–29, Jul. 2018.

[7] A. T. Sherman, L. Oliva, D. DeLatte, E. Golaszewski, M. Neary, K. Patsourakos, D. Phatak, T. Scheponik, G. Herman, and J. Thompson, "Creating a cybersecurity concept inventory: A status report on the cats project," *National Cyber Summit*, pp. 1–5, Jun. 2017.

[8] A. T. Sherman, L. Oliva, E. Golaszewski, D. Phatak, T. Scheponik, G. Herman, D. S. Choi, S. Offenberger, P. Peterson, J. Dykstra, G. Bard, A. Chattopadhyay, F. Sharevski, R. Verma, and R. Vrecenar, "The cats hackathon: Creating and refining test items for cybersecurity concept inventories," in *IEEE Security and Privacy*, 2019.

[9] R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *American Journal of Physics*, vol. 66, pp. 64–74, Jan. 1998.

[10] D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," *The Physics Teacher*, vol. 30, pp. 141–158, Mar. 1992.

[11] T. Litzinger, P. Van Meter, C. Firetto, L. J. Passmore, C. B. Masters, S. R. Turns, G. L. Gray, F. Costanzo, and S. Zappe, "A cognitive study of problem solving in statics," *Journal of Engineering Education*, vol. 99, pp. 337–353, Oct. 2010.

[12] D. Evans, G. Gray, S. Krause, J. Martin, C. Midkiff, B. Notaros, M. Pavelich, D. Rancour, T. Reed, P. S. Steif, R. Streveler, and K. Wage, "Progress on concept inventory assessment tools," *IEEE Frontiers in Education Conference (FIE)*, pp. T4G–1–T4G–8, Nov. 2003.

[13] K. Douglas and S. Purzer, "Validity: Meaning and relevancy in assessment for engineering education research: Assessment validity for engineering education research," *Journal of Engineering Education*, vol. 104, no. 2, pp. 108–118, Apr. 2015.

[14] J. Libarkin, "Concept inventories in higher education science," *National Research Council Promising Practices in Undergraduate STEM Education Workshop*, Oct. 2008.

[15] National Research Council, Division of Behavioral and Social Sciences and Education, Board on Testing and Assessment, Center for Education, Committee on the Foundations of Assessment, *Knowing What Students Know: The Science and Design of Educational Assessment*, J. W. Pellegrino, N. Chudowsky, and R. Glaser, Eds. Washington, DC: The National Academies Press, 2001.

[16] B. B. Brown, *Delphi Process A Methodology Used for the Elicitation of Opinions of Experts*. The RAND Corporation, 1968.

[17] G. Herman, C. Zilles, and M. C. Loui, "A psychometric evaluation of the digital logic concept inventory," *Computer Science Education*, vol. 24, pp. 277–303, Oct. 2014.

[18] N. Jorion, B. Gane, K. James, L. Schroeder, L. V. DiBello, and J. Pellegrino, "An analytic framework for evaluating the validity of concept inventory claims," *Journal of Engineering Education*, vol. 104, pp. 454–496, Oct. 2015.

[19] J. Ryan and F. Brockmann, *A Practitioners Introduction to Equating with Primers on Classical Test Theory and Item Response Theory*. Distributed by ERIC Clearinghouse, Jun. 2009.

[20] J. Cappelleri, J. Lundy, and R. Hays, "Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures," *Clinical Therapeutics*, vol. 36, pp. 648–662, May. 2014.

[21] L. J. Cronbach, "Coefficient alpha and internal structure of tests," *Psychometrika*, vol. 16, pp. 297–334, Sept. 1951.

[22] P. Panayides, "Coefficient alpha: Interpret with caution," *Europe's Journal of Psychology*, vol. 9, no. 4, pp. 688–696, Nov. 2013.

[23] S. Testa, A. Toscano, and R. Rosato, "Distractor efficiency in an item pool for a statistics classroom exam: Assessing its relation with item cognitive level classified according to blooms taxonomy," *Frontiers in Psychology*, vol. 9, no. 1585, pp. 1–12, Aug. 2018.

[24] D. DiBattista, J.-A. Sinnige-Egger, and G. Fortuna, "The none of the above option in multiple-choice testing: An experimental study," *The Journal of Experimental Education*, vol. 82, no. 2, pp. 168–183, 2014.

APPENDIX A ITEM DETAILS

Table A.1: Information About CCI Items

Name	ID	Estimated Difficulty	Concept	Topic	Scenario
Q1	A1-1	Medium	T	MAC	A1
Q2	A2-1	Easy	G	MAC	A2
Q3	A2-4	Medium	T	Non-Repudiation	A2
Q4	A3-3	Medium	D	Input Validation	A3
Q5	A4-1	Medium	G	Network Design	A4
Q6	A4-2	Medium	D	Network Design	A4
Q7	A4-3	Medium	V	Network Design	A4
Q8	B1-2	Medium	C	Replay Attack	B1
Q9	B1-3	Medium	C	Integrity	B1
Q10	B2-1	Medium	C	Physical Attack	B2
Q11	B2-2	Medium	T	Insider Threat	B2
Q12	B3-1	Easy	V	Security Theater	B3
Q13	B4-1	Hard	D	PKC	B4
Q14	B4-2	Medium	G	Replay Attack	B4
Q15	B4-3	Medium	V	Authentication	B4
Q16	B4-4	Medium	V	PKC	B4
Q17	C1-1	Easy	T	Authentication	C1
Q18	C2-1	Medium	G	Authorization	C2
Q19	C3a-1	Easy	D	Encryption	C3
Q20	C3b-1	Hard	V	Linkage	C3
Q21	C4-1	Easy	T	Social Engineering	C4
Q22	C4-2	Medium	C	Biometric Authentication	C4
Q23	D1-1	Easy	D	Network Isolation	D1
Q24	T1-1	Medium	G	Selecting Targets	T1
Q25	Z2-1	Hard	C	Protocols	Z2

APPENDIX B FULL ASSESSMENT

This appendix includes the full CCI as given to the students in the pilot test. The instructions and introduction are as seen by the students.

Instructions

Thank you for taking this draft Cybesecurity Concept Inventory (CCI), which was developed as part of the CATS Project.

Your participation will help us improve the assessment. You have 50 minutes to answer 25 questions.

Each test item has a scenario and a question (stem). Some questions share a common scenario, but each question should be answered independently. For each question, choose exactly one answer--the single best alternative from among the five given.

We also invite you to answer an additional seven optional questions in 22 minutes in the auxillary exam.

Informed consent

You must be of 18 years or older to participate in this survey.

The purpose of this study is to provide infrastructure for rigorous evidence-based improvement of cyber security education by developing the first Cyber Security Assessment Tools (CATs) targeted at measuring the quality of instruction, which can help universities better prepare the substantial number of cyber security professionals needed in America. You are being asked to volunteer because you are in or have completed a cybersecurity course. It will take up to 60 minutes to complete this survey.

We are developing the Cybersecurity Concept Inventory (CCI) in hopes of creating a nationally recognized, research-based assessment tool for measuring how well students use adversarial thinking after a first course in cybersecurity. The CCI will help us identify best practices for preparing future cybersecurity professionals. To verify whether the CCI will meet these goals, we are asking you to answer each of the questions on the survey.

There are no known risks involved in completing the survey. There are no tangible benefits for completing the survey, but you will have access to all results and project activities on the web site - http://www.cisa.umbc.edu/cats/index.html.

Participation is entirely voluntary; you may withdraw from participation at any time. If you withdraw from this research study, you will not be penalized in any way for deciding to stop participating.

All data obtained will be confidential. Results from this research will be published in academic conferences and journals, but no personally identifying information will ever be shared.

This study has been reviewed and approved by the UMBC Institutional Review Board (IRB). A representative of that Board, from the Office for Research Protections and Compliance, is available to discuss the review process or my rights as a research participant. Contact information of the Office is (410) 455-2737 or compliance@umbc.edu.

Demographics

This information will be anonymous from the user.

School

Professor

Year in School

- ○ Freshman
- ○ Sophomore
- ○ Junior
- ○ Senior
- ○ Graduate

Question 1

Scenario A1. A company delivers packages to customers using drones. The company's command center controls the drones by exchanging messages with them. The company's command center authenticates each message with a keyed message authentication code (MAC), using a key that is known by the command center and installed in each drone at initialization. The command center stores this key encrypted in a database.

Choose the post promising action for a malicious adversary to masquerade as the command center:

- Jam the command center’s signals and replace them.
- Capture a drone and extract its secret key.
- Exploit a vulnerability in the command center's firewall to access the database that contains the authentication key.
- Bribe an employee to give you the proprietary source code of the drone.
- Try to find messages with MACs that collide with the MACs of legitimate messages.

Definitions:

to masquerade: To pretend to be someone else.

Question 2

Scenario A2. Alice wants to send a file to Bob over an Internet connection.

Alice sends to Bob the file and a tag. The tag is the output of a message authentication code applied to the file and a key known only by Alice and Bob. Charlie is a malicious actor monitoring the connection between Alice and Bob.

Choose the action by Charlie that this tag mitigates:

- Pretend to be Alice by resending the file and the tag.
- Collude with Bob to forge a file sent by Alice.
- Recover the file contents.
- Change bits of the file in transit.
- Prevent Alice from completing the file transfer.

Question 3

Scenario A2. Alice wants to send a file to Bob over an Internet connection.

Bob receives a file digitally signed with Alice's private (signature) key, using a secure digital signature algorithm. The file specifies an electronic order to purchase a large number of shares for a new public offering. Contrary to expectation, the value of the stock plummets. Following this incident, Alice denies having signed the purchase order, pointing out that Charlie has been caught forging her signature.

Choose the most likely explanation for how Charlie forged Alice's signature:

- Copied Alice's digital signature from an older electronic purchase order.
- Mathematically analyzed Alice's signature to deduce her private key.
- Changed bits in Alice's signature to sign another electronic document.
- Received Alice's private key from Alice.
- Created a new document producing the same digital signature.

Question 4

Scenario A3. When a user Mike O'Brien registered a new account for an online shopping site, he was required to provide his username, address, first and last name, and a password. Immediately after Mike submitted his request, you -- as the security engineer -- observe a database input error message in the logs.

Choose the best defense to protect against possible security problems suggested by this error:

- Implement the system in a more secure programming language.
- Sanitize input at the server side.
- More thoroughly test the software before deploying it.
- Encrypt and authenticate all messages between the client and the server.
- Require all characters input by the user to be from a restricted set of characters.

Question 5

Scenario A4. An enterprise with highly sensitive data needs to be able to retrieve information from the Internet. To support this requirement while protecting its sensitive data, the enterprise partitions its internal computer network into three segments: Public, Quarantine, and Private. In this system, data should flow only from Public to Quarantine, and from Quarantine to Private.

Choose the most important security objective for this system:

- Maintain integrity of sensitive data.
- Detect intrusions.
- Block malicious traffic with a firewall.
- Restrict access to only trusted users.
- Prevent data exfiltration.

Definitions:

quarantine: A place of isolation in which potentially infectious items are placed and checked.

Question 6

Scenario A4. An enterprise with highly sensitive data needs to be able to retrieve information from the Internet. To support this requirement while protecting its sensitive data, the enterprise partitions its internal computer network into three segments: Public, Quarantine, and Private. In this system, data should flow only from Public to Quarantine, and from Quarantine to Private.

Choose the most effective method to prevent unwanted data flows:

- Authenticate all data flows.
- Restrict access to authorized users only.
- Encrypt all data flows.
- Use only one-way physical connections between the segments.
- Install software firewalls between the segments.

Definitions:

quarantine: A place of isolation in which potentially infectious items are placed and checked.

Question 7

Scenario A4. An enterprise with highly sensitive data needs to be able to retrieve information from the Internet. To support this requirement while protecting its sensitive data, the enterprise partitions its internal computer network into three segments: Public, Quarantine, and Private. In this system, data should flow only from Public to Quarantine, and from Quarantine to Private.

A malicious file was discovered on the Private segment. Choose the most likely cause of this failure:

- A user on Public visited a malicious website.
- The malicious file was not identified as bad in quarantine.
- The system administrator failed to update software on Private.
- A former employee who knew the network architecture created the file.
- A firewall in front of Private was misconfigured.

Definitions:

quarantine: A place of isolation in which potentially infectious items are placed and checked.

Question 8

Scenario B1. A bank offers online banking services. To connect to these services from her home computer, the user searches for the bank's name and follows the first link returned by the search. She logs into the website by entering her username and password. She then performs several banking transactions.

Alice logged on to the bank's website to make a payment of one thousand dollars to Bob. The next day, she discovers that an additional one thousand dollars were transferred to Bob's account. Bob had intercepted the encrypted traffic between Alice and the bank's website.

Choose the mostly likely explanation for how Bob used the intercepted traffic to transfer the additional one thousand dollars:

- Hijacked Alice's connection by extracting session data.
- Decrypted the traffic to discover the login credentials.
- Sent a copy of the encrypted traffic to the bank's website.
- Reverse engineered the banking protocol.
- Tricked the bank's online customer service.

Question 9

Scenario B1. A bank offers online banking services. To connect to these services from her home computer, the user searches for the bank's name and follows the first link returned by the search. She logs into the website by entering her username and password. She then performs several banking transactions.

Mary logged onto the bank's website and requested a transfer of two thousand dollars. Subsequently, she discovered that, instead, ten thousand dollars were transferred. A criminal was able to modify the transaction amount by modifying the traffic.

Choose the vulnerability that most likely explains this event:

- The transmitted data was not protected by an error-correcting code.
- The data format was publicly revealed.
- The protocol lacked two-factor authentication.
- The transaction was not protected by a signed hash.
- The bank's firewall was configured incorrectly.

Question 10

Scenario B2. While Mary is traveling she decides to do some shopping online. She connects from a computer in a hotel business office that uses a wired network.

A fellow hotel guest wants to steal Mary's credit card information. The hotel's IT staff monitors machines in the business office using anti-malware and network intrusion-detection software.

Choose the attack that is least likely to be detected by the IT staff:

- Perform a man-in-the-middle attack on the hotel's local network to masquerade as the online shopping service.
- Establish screen sharing with Mary's machine from another machine in the business office.
- Plant a key-stroke logging device between Mary's keyboard and machine.
- Monitor traffic on the hotel's local network to observe packets sent by Mary's machine.
- Insert a USB drive into the machine that automatically installs custom software.

Definitions:

masquerade: To pretend to be someone else.

Question 11

Scenario B2. While Mary is traveling she decides to do some shopping online. She connects from a computer in a hotel business office that uses a wired network.

The hotel manager is concerned that for the past year, many hotel guests have reported credit card thefts after visiting the business center. Repeated attempts to solve this problem, such as reimaging machines, purchasing new software and hardware, and encrypting network traffic have all failed.
Choose the most likely explanation for these attacks:

- The purchased hardware includes backdoors implemented by the manufacturer.
- Hotel guests frequently visit dubious websites from machines in the business center.
- Attackers have defeated the security of protocols protecting guest-to-server communications.
- Attackers are monitoring traffic on the business center's local network.
- A member of the IT staff is involved in the thefts.

Question 12

Scenario B3. A soft drink company electronically stores the secret formula for its popular drink on a computer that is disconnected from all networks. A competitor wants to the learn the secret formula.

Choose the action that is LEAST useful at preventing the competitor from learning this formula.

- Encrypt the sensitive data.
- Change passwords frequently.
- Require two people to access the data.
- Perform thorough background checks on all employees.
- Protect the facility with guards and fences.

Question 13

Scenario B4. To comply with the terms of a nuclear test ban treaty, Country A would like to implant a seismic sensor under Country B's soil to monitor underground weapons testing. Country A fears that B will try to falsify the signals of the sensor, and Country B fears that A will try to exfiltrate spy information embedded in the seismic data. Neither party trusts the other.
Requirements of the system are:

1. *Country A wants assurance that the seismic data it receives came from its sensor and were not modified.*
2. *Country B wants to be able to monitor the signals transmitted from the sensor in real time. It also wants assurance that the signals were not modified.*
3. *The design should be fair to both parties.*

A device is placed in a deep, narrow hole underground in Country B and certain cryptographic keys are distributed to the device, Country A, and Country B. The device comprises a seismic sensor and cryptographic hardware for processing its signals before they are broadcast to a satellite.

To best satisfy the system requirements, choose what type of cipher the device should apply to each message:

- Symmetric cipher separately applying two keys, one known only by A and the device, and one known only by B and the device.
- Symmetric cipher applying a key known by A, B, and the device.
- Asymmetric cipher (public-key cryptography) applying a private key known only by the device. The corresponding public key is known by A, B, and the device.
- Asymmetric cipher (public-key cryptography) applying a private key known only by the device and A, with the corresponding public key known by A, B, and the device.
- Asymmetric cipher (public-key cryptography) applying a public key known by A, B, and the device. The corresponding private key is known only by the device.

Question 14

Scenario B4. To comply with the terms of a nuclear test ban treaty, Country A would like to implant a seismic sensor under Country B's soil to monitor underground weapons testing. Country A fears that B will try to falsify the signals of the sensor, and Country B fears that A will try to exfiltrate spy information embedded in the seismic data. Neither party trusts the other.
Requirements of the system are:

1. *Country A wants assurance that the seismic data it receives came from its sensor and were not modified.*
2. *Country B wants to be able to monitor the signals transmitted from the sensor in real time. It also wants assurance that the signals were not modified.*
3. *The design should be fair to both parties.*

Data are chunked into messages, each with a unique sequential message number.

Choose the security goal that is supported by including unique message numbers:

- Protect against replay attacks.
- Prevent any messages from producing equal hash values.
- Facilitate the retransmission of lost or garbled messages: the recipient can specify by message number which messages need to be retransmitted.
- Protect against man-in-the-middle attacks.
- Provide ordering information necessary to verify authenticity of the message.

Question 15

Scenario B4. To comply with the terms of a nuclear test ban treaty, Country A would like to implant a seismic sensor under Country B's soil to monitor underground weapons testing. Country A fears that B will try to falsify the signals of the sensor, and Country B fears that A will try to exfiltrate spy information embedded in the seismic data. Neither party trusts the other.
Requirements of the system are:

1. *Country A wants assurance that the seismic data it receives came from its sensor and were not modified.*
2. *Country B wants to be able to monitor the signals transmitted from the sensor in real time. It also wants assurance that the signals were not modified.*
3. *The design should be fair to both parties.*

Consider a design in which the device applies a keyed cryptographic hash function to each message using a key distributed only to the device, Country A, and Country B.

To enable both countries to authenticate each message and satisfy all system requirements, choose what the device should output:

- The message together with a hash of the following: message and current time.
- The key together with a hash of the message.
- The message together with a hash of the message.
- A hash of the message.
- This design cannot satisfy the system requirements.

Question 16

Scenario B4. To comply with the terms of a nuclear test ban treaty, Country A would like to implant a seismic sensor under Country B's soil to monitor underground weapons testing. Country A fears that B will try to falsify the signals of the sensor, and Country B fears that A will try to exfiltrate spy information embedded in the seismic data. Neither party trusts the other.
Requirements of the system are:

1. *Country A wants assurance that the seismic data it receives came from its sensor and were not modified.*
2. *Country B wants to be able to monitor the signals transmitted from the sensor in real time. It also wants assurance that the signals were not modified.*
3. *The design should be fair to both parties.*

The cryptographic hardware encrypts with a widely-deployed symmetric cipher with key k known by Country A, Country B, and the device.

Choose the the most serious limitation of this authentication strategy:

- Symmetric ciphers provide confidentiality, not authentication.
- Anyone who knows k can forge a message.
- It is difficult to distribute a new key k.
- Using encryption makes it more difficult to recover from transmission errors.
- If a party misplaces k, they cannot authenticate any message.

Question 17

Scenario C1. 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."

Upon receiving Alice's message, choose what action Bob should take:

- Check that the source address in the received email exactly matches Alice's corporate e-mail.
- Ask Alice via email for the address of her hotel, and then send the specifications by courier.
- Send the encrypted specifications to Alice via email and then text her the decryption key.
- Verify that Alice sent the email requesting the technical specifications.
- Reply to Alice's email, attaching the specifications together with a cryptographic hash.

Definitions:

courier: A special-purpose messenger.

Question 18

Scenario C2. A security company is designing a precinct voting system to enable members of the military to vote in their hometown elections by casting a ballot at their overseas military base. The system must provide, among other properties, voter authentication, ballot confidentiality, integrity of marked ballots, and assured operations. When a voter checks in, a pollworker records the voter's name on the voter check-in list and verifies each of the following:

1. *The voter's appearance resembles the image on an acceptable photo identification card.*
2. *The voter's name is on the voter registration list.*

Identify the primary purpose of checking if the voter's name is on the voter registration list:

- Be able to verify later that the person cast a ballot.
- Prevent criminals from casting votes on behalf of deceased people.
- Determine whether the person is eligible to vote at this location.
- Ensure that the person votes only once.
- Verify the person is who they claim to be.

Definitions:

precinct voting: Voting that takes place in an enclosed area which is supervised and provides a considerable degree of security and ballot privacy.

Question 19

Scenario C3a. Alice logs into a server by sending her username, password, and a timestamp to the server over the Internet. Eve listens to the communications.

Eve wishes to log into Alice's account at a later time. Choose the modification to the login protocol that best deters Eve from later logging in to Alice’s account:

- Instead of sending the username, password, and timestamp over the Internet, send them in a text message from Alice's smartphone.
- In a second step, prompt Alice for additional personal information, such as her mother's maiden name.
- Replace the username and password pair with Alice's fingerprint.
- Encrypt the username, password, and timestamp.
- Include a use-once, randomly generated value to prevent replay attacks.

Question 20

Scenario C3b. Alice is logging onto a server from her laptop. She sends her username and password to the server over the Internet. The server then instructs the security computer to send a challenge to Alice's cell phone, which she answers using a text message. For example, the challenge is the name of Alice's pet, and the response is "Skippy". Upon deeming the response valid, the security computer signals the server to accept Alice's login request.

Eve has obtained Alice's username and password and has positioned herself in the middle of Alice's Internet communication with the server. While this allows Eve to block and inject messages between Alice and the server, Eve is not able to do the same between Alice's cell phone and the security computer.

Choose the vulnerability that will most likely allow Eve to log in as Alice:

- Alice's keystrokes on her laptop are being monitored by malware propagated by Eve.
- The challenge sent by the security computer does not reference Alice's login request.
- Eve can send messages to Alice that appear to have originated from the server.
- Alice's firewall is misconfigured, allowing Eve to monitor all communications with the server.
- The answer to Alice's security challenge is easy to guess.

Definitions:

masquerade: To pretend to be someone else.

Question 21

Scenario C4. Alice runs a top-secret government facility where she has hidden a USB stick, with critical information, under a floor tile in her workspace. The facility is secured by guards, 24/7 surveillance, fences, electronically locked doors, sensors, alarms, and windows that cannot be opened. To gain entrance to the facility, all employees must present a cryptographically hardened ID card to guards at a security checkpoint. All of the computer networks in the facility use state-of-the-art computer security practices and are actively monitored.

Alice hires Mark (an independent penetration tester) to exfiltrate the data on the USB stick hidden in her workspace.

Choose the strategy that best avoids detection while effectively exfiltrating the data:

- Compromise the facility's network and add Mark as an authorized guest.
- Convince an authorized employee to remove the USB stick and give it to Mark.
- Unlock electronically-locked doors using malware.
- Climb over the perimeter fence at night and sneak into Alice's workspace.
- Fabricate a fake ID to fool the guards at the security checkpoint.

Definitions:

24/7: Twenty-four hours a day, seven days a week.

Question 22

Scenario C4. Alice runs a top-secret government facility where she has hidden a USB stick, with critical information, under a floor tile in her workspace. The facility is secured by guards, 24/7 surveillance, fences, electronically locked doors, sensors, alarms, and windows that cannot be opened. To gain entrance to the facility, all employees must present a cryptographically hardened ID card to guards at a security checkpoint. All of the computer networks in the facility use state-of-the-art computer security practices and are actively monitored.

Alice's workspace is protected by an electronically-locked door featuring a tamper-resistant fingerprint scanner. To unlock the door, the electronic door lock scans a fingerprint and checks it against an encrypted fingerprint stored in a database. The encryption keys are secret. Additionally, the door is monitored continuously by a camera whose output is observed by a security guard.

Choose the action that best exploits this system:

- Try many possible fingerprints using a device that quickly outputs randomly-chosen fingerprints to the scanner.
- Use a screwdriver or other physical tools to expose the electronics of the lock to unlock the door by gaining access to the lock control wire or physical door latch.
- In the database, replace Alice's encrypted fingerprint with the attacker's encrypted fingerprint.
- Lift fingerprints from objects touched by Alice to recreate her fingerprint.
- Retrieve Alice's fingerprint by decrypting the encrypted copy stored in the database.

Definitions:

24/7: Twenty-four hours a day, seven days a week.

Question 23

Scenario D1. A law firm stores sensitive client records in a database on their local network.

Choose the action that is the MOST likely to prevent an opposing law firm from reading the records:

- Require fingerprint scans to access the law offices.
- Disconnect their local network from the Internet
- Use only trusted vendor software.
- Protect the network with a state-of-the-art firewall and intrusion-detection system.
- Secure the law offices 24/7 with strong locks and security cameras.

Definitions:

24/7: Twenty-four hours a day, seven days a week.

Question 24

Scenario T1: Mike is a penetration tester performing a penetration test of ACME Corporation, a leading producer of automobiles. Mike has been given a wired connection to ACME's local area network (LAN).

Choose the node that Mike should first compromise to have the highest likelihood of success in remaining undetected and understanding how ACME Corporation operates:

- Server containing data about automobile prototypes.
- Database storing human resources information.
- Firewall between ACME's LAN and the Internet.
- Server that stores logs of network connections for the LAN.
- Server responsible for authenticating all users on the LAN.

Question 25

Scenario Z2. Bob wishes to send a sensitive document D to Alice. To accomplish this goal, they wish to establish a shared session key k for symmetric encryption. For asymmetric encryption, Alice has her own secret key sA and Bob's authenticated public key pB. Similarly, Bob has his own secret key sB and Alice's authenticated public key pA. To encrypt and decrypt, Alice and Bob use a strong symmetric cipher E and a strong asymmetric cipher R. Let E(k, D) denote symmetric encryption of D with key k, and let R(sA, D) denote asymmetric encryption of D with key sA.

Alice and Bob agree on the following protocol:

1. Alice generates a session key k at random.
2. Alice sends c = R(sA, k) to Bob.
3. Bob receives c and decrypts c by computing k = R(pA, c).
4. Bob sends E(k, D) to Alice.

where
pA = Alice's public key,
sA = Alice's secret key,
pB = Bob's public key, and
sB = Bob's secret key.

Choose the most fundamental flaw of this protocol:

- A passive eavesdropper can read the document D.
- Bob cannot decrypt c.
- An adversary can modify bits of c in transit.
- It is computationally expensive to compute several encryptions/decryptions.
- The session key k is not authenticated by Alice.

Definitions:

masquerade: To pretend to be someone else.