Psychometric Evaluation of the Cybersecurity Curriculum Assessment

Alan T. Sherman
Cyber Defense Lab, University of Maryland, Baltimore County

SIGCSE, Toronto, March 16, 2023

Geoffrey L Herman, Shan Huang, Peter A. Peterson, Enis Golaszewski, Linda Oliva

UMD DULUTH

UMBC

This work was supported in part by the U.S. Department of Defense and the National Science Foundation

Main results

We developed and validated two concept inventories for cybersecurity and found they are valid and reliable.

- What is the CCA?
- How did we develop it?
- Example test item
- Expert review
- Statistical analysis using Classical Test Theory (CTT) & Item Response Theory (IRT)

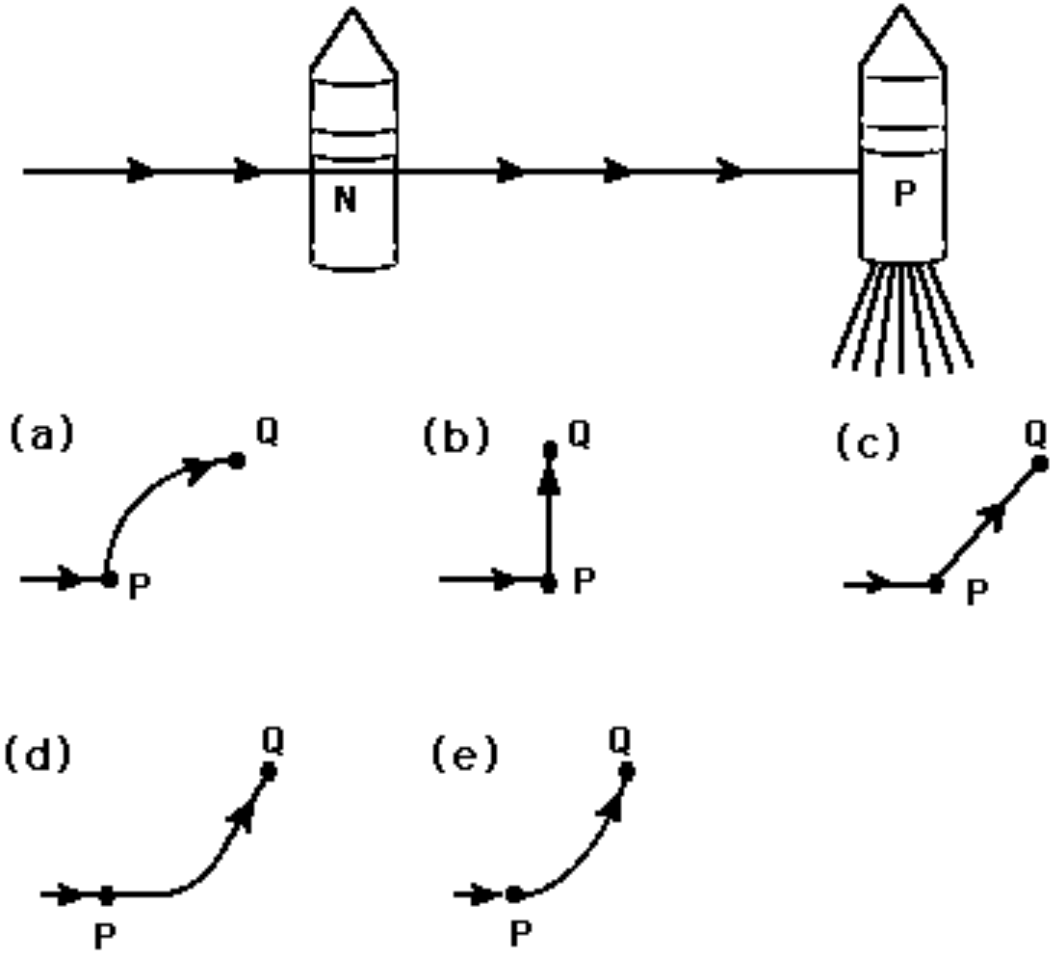
We need a workforce that deeply understands cybersecurity

What is a Concept Inventory?

Multiple-choice

Conceptual understanding

Core concept(s)



What Cybersecurity Assessment Tools (CATs) are we developing?

1

CCI

Cybersecurity Concept Inventory

- After first course

2

CCA

Cybersecurity Curriculum Assessment

- After full curriculum

How did we develop the CATs?

Identify Core Concepts

- Delphi process
- Expert review/co-development

Model Cognition

- Open-Ended interviews
- Draft CATS questions

Statistical Validation

- Administer prototypes
- Psychometric analysis
- Re-administer with bigger *n*

What concepts are on the CCI and CCA?

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

- CIA = Confidentiality, Integrity, Availability

How did we develop the CATs?

Identify Core Concepts	• Delphi process • Expert review/co-development
Model Cognition	• Open-Ended interviews • Draft CATS questions
Statistical Validation	• Administer prototypes • Psychometric analysis • Re-administer with bigger *n*

Open-ended questions developed

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.

What are potential security problems suggested by this error?

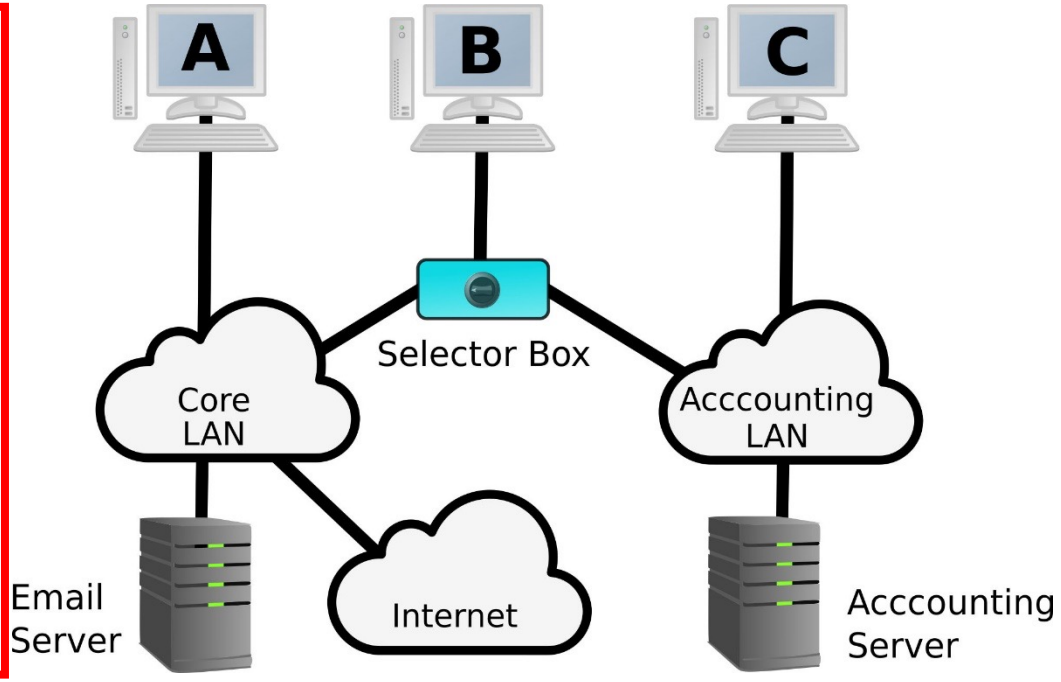
How would you defend against them?

4 themes in student misconceptions

1. Conflation (Threat and risk)
2. Overgeneralization (Encryption secures everything)
3. Bias (e.g., Cyber safer than physical, Trust the user)
4. Limiting adversaries (Only consider theft)

CCA items comprise a scenario, stem, and choices

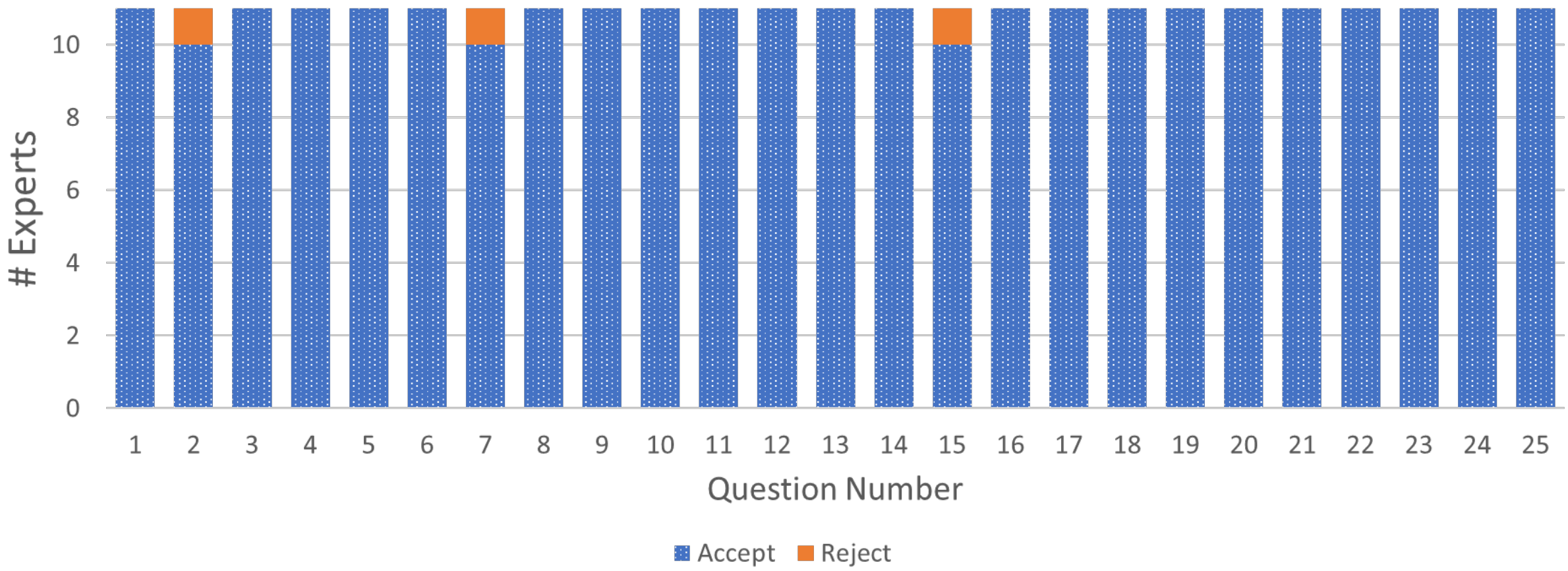
A company has two internal Local Area Networks (LANs): a core LAN connected to an email server and the Internet, and an accounting LAN connected to the corporate accounting server (which is not connected to the Internet). Each desktop computer has one network interface card. Computers A and C are connected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



Choose the action that this design best prevents:

(a) Emailing accounting data.
(b) Infecting the accounting LAN with malware.
(c) Employees accessing the accounting server from home.
(d) User of Computer B accessing the accounting LAN without authorization.
(e) Computer A communicating with Computer B.

A majority of experts approve all CCA items



How did we develop the CATs?

Identify Core Concepts
- Delphi process
- Expert review/co-development

Model Cognition
- Open-Ended interviews
- Draft CATS questions

Statistical Validation
- Administer prototypes
  278 students, 7 institutions
- Psychometric analysis

Psychometric Frameworks

- Classical Test Theory
- Item Response Theory

What do we want to measure?

reliability*:* Is the score repeatable?

difficulty: How hard is each question?

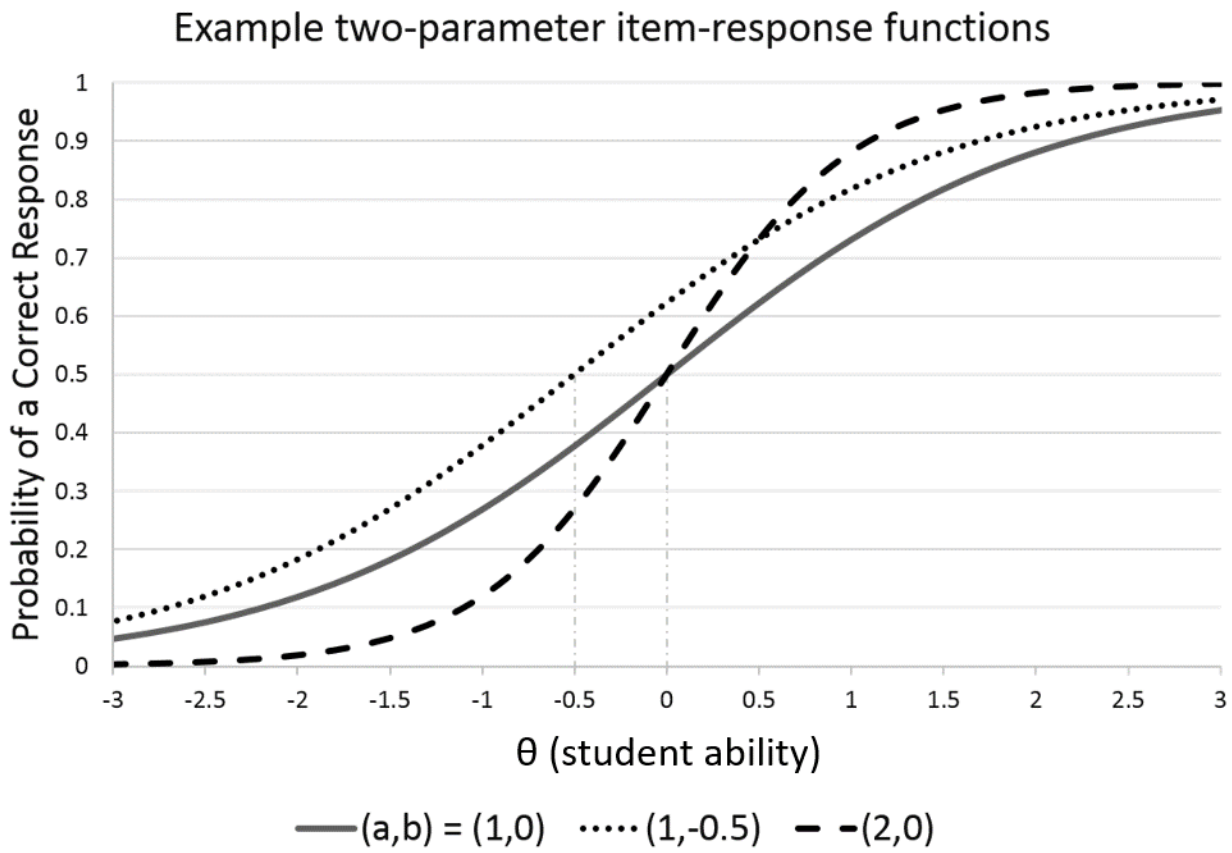
discrimination: Do students of lower and higher skill levels perform differently?

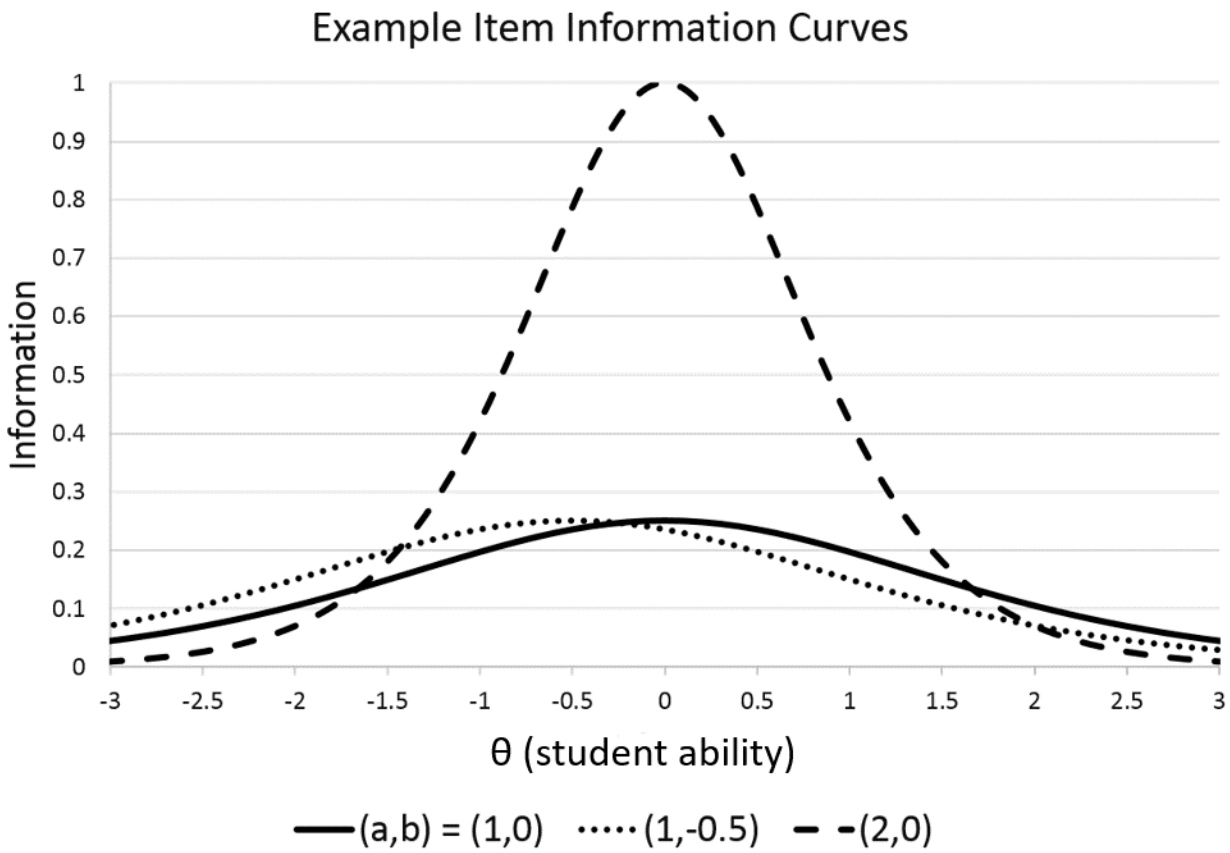
information: How precise is the score for students at different skill levels?

IRT: Two-parameter Logistic (2PL) Model

$$P\{X_{n,i} = 1\} = \frac{1}{1 + e^{-a_i(\theta_n - d_i)}}$$

- θ_n is the ability of student n (normalized)
- d_i is the difficulty of item i
- a_i is the discrimination of item i

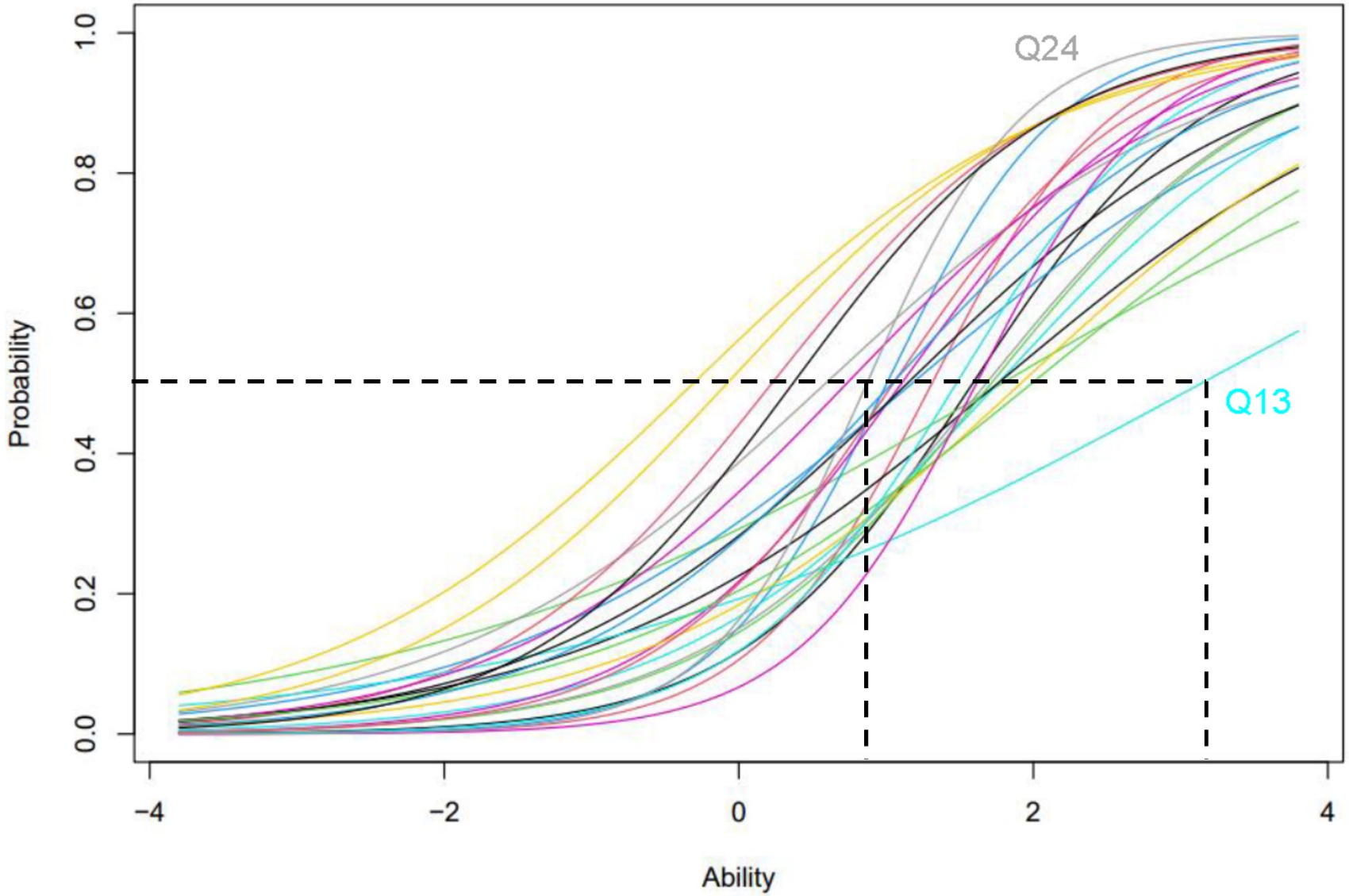




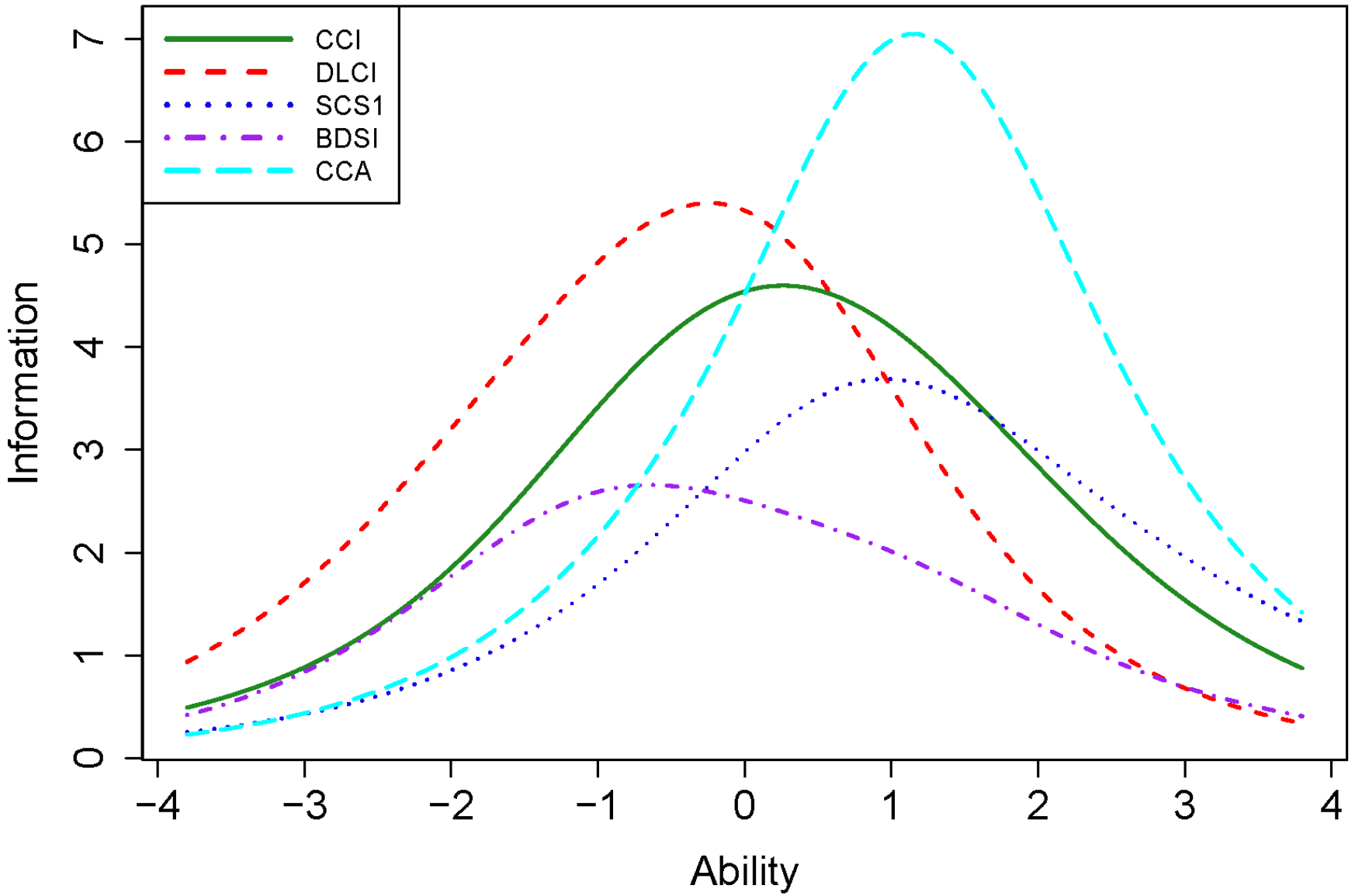
Important Summary Statistics

Measurement	CCA	CCI	DLCI	SCS1	BDSI
Cronbach's α	0.83	0.78	0.80	0.70	0.68
Min Difficulty	-0.31	-2.00	-1.84	0.08	-3.03
Max Difficulty	3.14	2.04	0.55	5.07	1.255
Min Discrimination	0.46	0.37	0.28	0.49	0.33
Max Discrimination	1.99	1.47	1.68	1.53	2.03

IRT: Two-parameter Logistic (2PL) Model



Item Information Curves for CS Assessments



Future Work

- More item-level analysis of the CCA

- Large scale pedagogical testing at Military Academies

Our Contributions

- We performed a psychometic evaluation of the CCA completed by 193 students
- The CCA is reliable measurement of student knowledge
- The CCA is a difficult assessment that provides deeper information about high-performing students

Difficulty & Discrimination (CTT)

Item	Diff.	Disc.	Item	Diff.	Disc.
Q1	0.17	0.52	Q14	0.36	0.41
Q2	0.17	0.61	Q15	0.21	0.41
Q3	0.30	0.33	Q16	0.19	0.46
Q4	0.22	0.57	Q17	0.30	0.41
Q5	0.20	0.45	Q18	0.25	0.51
Q6	0.25	0.52	Q19	0.22	0.36
Q7	0.55	0.37	Q20	0.3	0.45
Q8	0.39	0.39	Q21	0.17	0.52
Q9	0.24	0.36	Q22	0.13	0.59
Q10	0.44	0.45	Q23	0.51	0.39
Q11	0.18	0.47	Q24	0.23	0.6
Q12	0.32	0.39	Q25	0.40	0.46
Q13	0.20	0.27			

Table 2: Difficulty and discrimination of each CCA item in Classical Test Theory

Difficulty & Discrimination (IRT)

Item	Diff. (b_i)	Disc. (a_i)	Item	Diff. (b_i)	Disc. (a_i)
Q1	1.59	1.27	Q14	0.73	0.87
Q2	1.31	1.63	Q15	1.91	0.78
Q3	1.79	0.50	Q16	1.68	1.03
Q4	1.00	1.71	Q17	1.14	0.81
Q5	1.76	0.91	Q18	1.05	1.24
Q6	1.11	1.16	Q19	1.99	0.68
Q7	-0.31	0.81	Q20	1.04	0.91
Q8	0.59	0.78	Q21	1.48	1.36
Q9	1.76	0.70	Q22	1.61	1.63
Q10	0.22	1.05	Q23	-0.08	0.90
Q11	1.71	1.04	Q24	0.87	1.89
Q12	1.18	0.71	Q25	0.37	1.13
Q13	3.14	0.46			

Table 3: Difficulty and discrimination of each CCA item using the 2PL Item Response Theory model.

Cronbach-Alpha: Removing Test Items

Item	$\Delta\alpha$	Item	$\Delta\alpha$	Item	$\Delta\alpha$
Q1	−0.009	Q10	−0.006	Q19	−0.003
Q2	−0.012	Q11	−0.007	Q20	−0.006
Q3	−0.001	Q12	−0.003	Q21	−0.009
Q4	−0.011	Q13	−0.000	Q22	−0.011
Q5	−0.006	Q14	−0.004	Q23	−0.003
Q6	−0.009	Q15	−0.005	Q24	−0.012
Q7	−0.002	Q16	−0.007	Q25	−0.006
Q8	−0.003	Q17	−0.004		
Q9	−0.002	Q18	−0.009		

Table 1: Change in Cronbach's α from removing each CCA item individually. Overall Cronbach's α is 0.83. Removing each item lowers reliability, suggesting that all items measure the same construct.