THE 8TH ANNUAL

CONFERENCE ON TEST SECURITY

OCTOBER 16-18, 2019 | MIAMI, FL

PRESENTED BY UNIVERSITY OF MIAMI

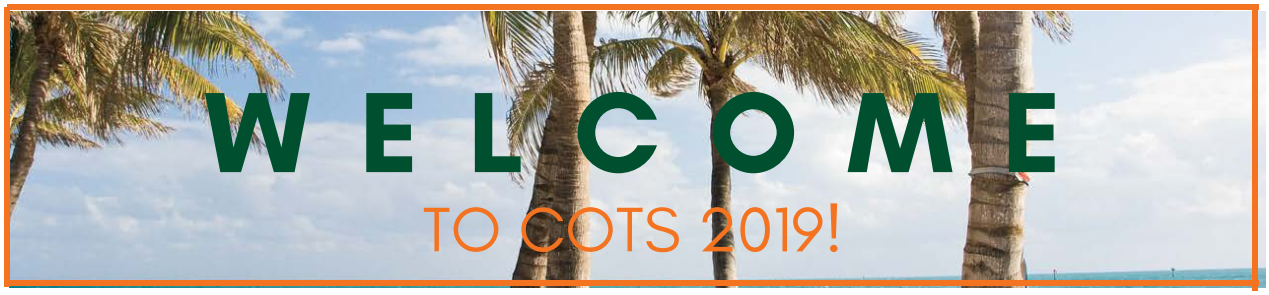
#COTS2019

Conference On Test Security 2019

THIS PROGRAM BELONGS TO:

PLEASE RETURN IF FOUND

FOLLOW US



Dear Colleague,

Welcome to the 8th Annual Conference on Test Security – the only event dedicated entirely to test security. It is our pleasure to have you with us in Miami, the cultural and economic center of the sunny South Florida. In this paradise of oceanfront cities, you will find many charms including beaches, tropical parks and gardens, a lively art scene, and sizzling cuisine.

When COTS was first started in 2012, it was called the Conference on the Statistical Detection of Potential Test Fraud, and focused primarily on statistics. In 2014, the scope of the conference broadened to encourage dialogue on all test security capabilities and enhancements. Now in its eighth year, COTS has rightfully gained a reputation for facilitating the best atmosphere encouraging open and honest test security discussions among professionals like yourselves. In the midst of cheating scandals, there is a waning trust in test scores and the validity of test scores is threatened. With over 30 sessions, and attendees from national and international organizations, we hope you'll find information and resources that are helpful to you and your program's needs.

If you're new to COTS, we hope that you'll join us for the Reception at the Pool Deck Lower on the evening of October 17th. Here, you'll find the opportunity to converse one-on-one with some of the brightest minds in testing and participate in our poster presentations. Dubbed as "the favorite event of COTS", you won't want to miss meeting with leaders in the field on this up-close and personal level.

Finally, a special thank you to our sponsors for making this year's conference possible. We appreciate their unyielding support. It is because of them that we can gather together and forward our joint goal to protect the validity of test results and brand integrity.

We hope that you enjoy your time in Miami, FL at the 2019 Conference on Test Security.

Warmest regards,

COTS 2019 Executive Committee

COTS
Conference On
Test Security

AN EXTRA SPECIAL THANK YOU TO OUR HOST

UNIVERSITY OF MIAMI

UNIVERSITY OF MIAMI

SCHOOL of EDUCATION & HUMAN DEVELOPMENT

A SPECIAL THANK YOU TO OUR CO-HOSTS

THANK YOU TO OUR FRIENDS

Conference Agenda

Wednesday, October 16th

9:00 AM - 6:30 PM	Registration & Information 3rd Floor Pre-Function Space
12:00 PM - 5:00 PM	Workshops Salon ABCD
5:00 PM - 6:00 PM	Meet & Greet 3rd Floor Pre-Function Space
5:00 PM - 6:30 PM	COTS Executive Committee Meeting Fisher Island
7:30 PM - 9:30 PM	The (Dis)Honesty Project Documentary Screening Regal Cinemas - South Beach

Thursday, October 17th

7:00 AM - 5:30 PM	Registration & Information 3rd Floor Pre-Function Space
7:00 AM - 7:45 AM	Continental Breakfast Salon F
7:45 AM - 8:15 AM	Welcome Dean Laura Kohn-Wood Salon E
8:15 AM - 9:30 AM	Opening Keynote Salon E

Conference Agenda

(Continued)

9:30 AM - 9:45 AM	Break
9:45 AM - 10:45 AM	Sessions Hibiscus Island, Salon AB, Salon C, and Salon D
10:45 AM - 11:00 AM	Break
11:00 AM - 12:00 PM	Sessions Hibiscus Island, Salon AB, Salon C, and Salon D
12:00 PM - 1:00 PM	Lunch Salon F
1:00 PM - 2:30 PM	Sessions Hibiscus Island, Salon AB, Salon C, and Salon D
2:30 PM - 3:00 PM	Extended Break
3:00 PM - 4:00 PM	Sessions Hibiscus Island, Salon AB, Salon C, and Salon D
4:00 PM - 4:15 PM	Break
4:15 PM - 5:15 PM	Sessions Hibiscus Island, Salon AB, Salon C, and Salon D
6:00 PM - 8:00 PM	Cocktails & Conversations Poster Presentations Pool Deck Lower (5th Floor)

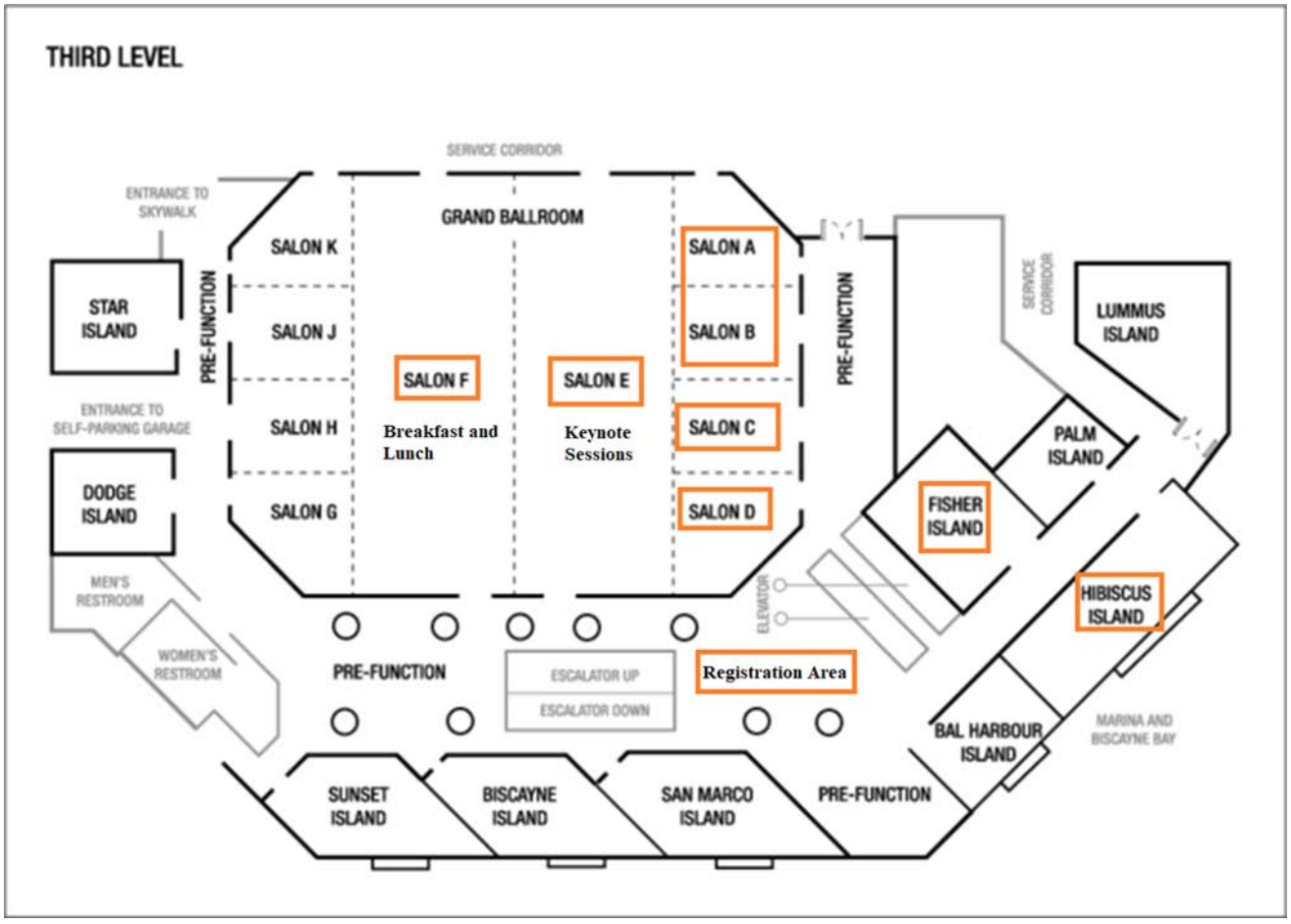
Conference Agenda

(Continued)

Friday, October 18th

7:00 AM - 12:00 PM	Registration & Information 3rd Floor Pre-Function Space
7:00 AM - 8:00 AM	Continental Breakfast Salon F
8:00 AM - 9:00 AM	Sessions Hibiscus Island, Salon AB, Salon C, and Salon D
9:00 AM - 9:15 AM	Break
9:15 AM - 10:30 AM	Closing Keynote Salon E
10:30 AM - 10:45 AM	Break
10:45 AM - 12:00 PM	Sessions Hibiscus Island, Salon AB, Salon C, and Salon D
12:00 PM - 1:00 PM	Lunch Salon F

Marriott Biscayne Bay
3rd Floor Map



* Registration desk will be close to elevators on the 3rd floor where all the conference events will be held.

* Please take the elevator to the 5th floor to access the Pool Deck Lower area for the Reception on th evening of Thursday, October 17.

WEDNESDAY, OCT 16
12:00 P.M. – 10:00 P.M.

12:00 - 1:30

SALON ABCD

EXAM SECURITY PLANNING FUNDAMENTALS

Jeff M. Marsh, Ascend Learning | Jennifer Geraets, The College Board | Rachel R. Watkins Schoenig, Cornerstone Strategies
Workshop

Developing an exam security framework can be a challenging process for any testing program. While it is helpful to understand what other programs have implemented, what makes sense for one program may or may not make sense for another. If there isn't a "recipe" for exam security, how can a program know if they're doing the right things?

This session will introduce participants to methods for analyzing their testing programs in order to determine how to right-size their test security framework. Using a high level of audience engagement, participants will learn what information is important to understand prior to analyzing risks for their programs. They will learn how to develop a plan to deter and detect test security incidents, how to be prepared to react when they do happen, and how to use those incidents to enhance security going forward for their program. Consideration will be given to factors such as delivery and proctoring models, risk tolerance, cost, and customer experience throughout this process. Finally, participants will leave with concrete suggestions for prioritizing future test security enhancements and working with others in their organizations to begin to implement those enhancements. Seasoned experts will guide this hands-on workshop that will empower you to confidently plan and operationalize your program's exam security framework. We will use scenarios and handouts to help attendees engage in the session and leave with practical tools for improving security.

1:45 - 3:15

SALON ABCD

BEEN THERE. HEARD THAT.

Camille Thompson, The College Board | Jennifer Semko, Baker McKenzie | Rachel R. Watkins Schoenig, Cornerstone Strategies
Workshop

Deciding how to respond to cheating attempts can be fraught with legal concerns. Examinees often raise a myriad of explanations for response similarities and unusual response times; third-party infringers may argue fair use and lack of substantial similarity for copyright purposes; and attorneys may raise various defenses and legal complaints, ranging from discrimination and due process violations to statistical weaknesses. What can a testing program anticipate when it comes to imposing exam security outcomes, and how can the program ensure it is in the best position possible to respond? As attorneys and as seasoned exam security experts, we will share the excuses we have heard, the defenses raised, and the court cases that provide guidance on imposing consequences associated with exam security incidents.

The goal is to empower testing programs to take appropriate action, without being overly fearful of threatened litigation. We will use interactive vignettes to engage the audience throughout the session. Participants will leave with concrete examples and case law to guide exam security outcomes and greater confidence in taking action to protect the integrity of their testing program and score results.

WEDNESDAY, OCT 16
12:00 P.M. – 10:00 P.M.

3:30 - 5:00 — *SALON ABCD*

THE VALUE OF DETERRENCE SOLUTIONS TO THE SECURITY OF TESTS

David Foster, Caveon Test Security
Workshop

Reducing test fraud requires carefully planned solutions. Those that detect the fraud and follow up with actions to mitigate it are the most common, and the ones we usually think of and use first. Next in line is prevention, which is usually accomplished by designing our tests and our test administration procedures carefully. But we don't often consider deterrent solutions. We do not often plan and implement solutions that are strictly intended to discourage or inhibit test fraud. These are psychological effects, primarily. we are attempting to convince and persuade cheaters and other ne'er-do-wells that test fraud isn't worth it. This session focuses on this third, less understood category of solutions, with the goal of raising awareness of its value, and providing instruction and direction in its use. Examples and audience participation demonstrations will illustrate how it works, and how it can be used more effectively.

5:00 - 6:00 — *3rd Floor Pre-Function Space*

MEET & GREET

Light refreshments will be served

5:00 - 6:30 — *FISHER ISLAND*

COTS EXECUTIVE COMMITTEE MEETING

7:30 - 9:30 — *REGAL SOUTH BEACH SCREEN X, IMAX, VIP Auditorium #2*

DOCUMENTARY SCREENING: THE (DIS)HONESTY PROJECT

Free round-trip transportation will be provided for conference attendees from the Marriott Biscayne Bay to the Regal South Beach Theaters. Three buses (each bus has a capacity of 55 seats) will depart at 7:00 PM from the Marriott Biscayne Bay and transport the conference attendees to the Regal South Beach Theaters. The documentary screening starts at 7:30 PM and it is about 90 minutes. There may be a brief discussion session following the screening. The buses will depart at 10:00 PM from the theater to the hotel. The auditorium has 163 seats.

Regal South Beach Screen X, IMAX, VIP: 1120 Lincoln Rd, Miami Beach, FL 33139

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

Conference On Test Security 2019

7:00 - 7:45

SALON F

CONTINENTAL BREAKFAST

7:45 - 8:15

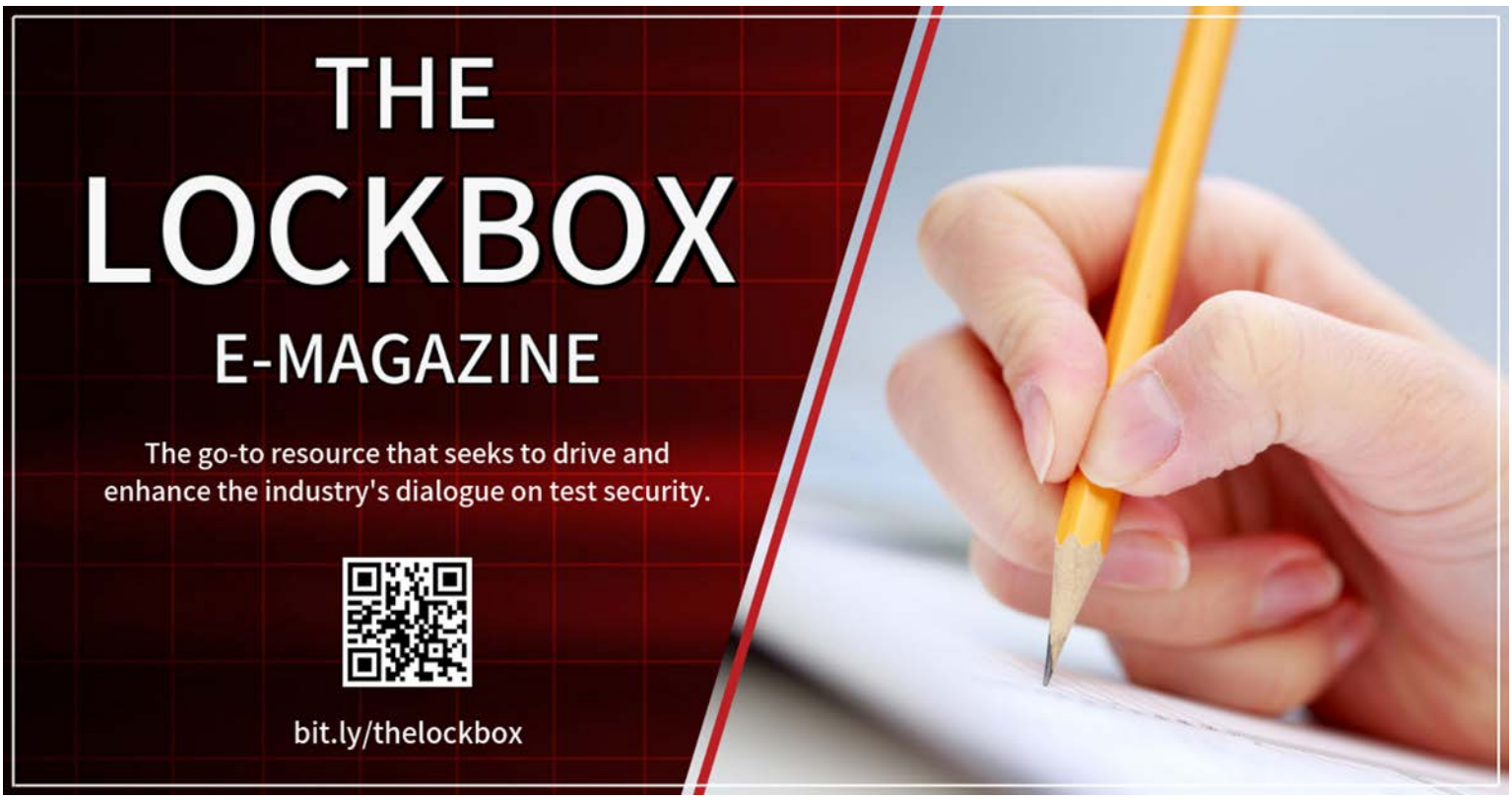
SALON E

WELCOME

DR. LAURA KOHN-WOOD, DEAN

SCHOOL OF EDUCATION AND HUMAN DEVELOPMENT

UNIVERSITY OF MIAMI



THURSDAY, OCT 17

OPENING KEYNOTE

8:15 - 9:30 AM | SALON E

COTS KEYNOTE DEBATES

Moderator: Rachel R. Watkins Schoenig, Cornerstone Strategies

Think there's nothing left to disagree about? Think again. The COTS debates are back by popular demand. Join seasoned professionals as they discuss some of your most pressing questions about the industry: the future of proctoring, critical threats, and the balance between security and validity. Bring your cell phones, as you will be asked to weigh in with your vote before and after each debate. Learn from the experts, shape and inform your own opinion, and gain from the wisdom of the crowd as we explore these important topics in assessment

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

9:30 - 9:45

BREAK

9:45 - 10:45

SALON AB

THE STRONGEST MODEL FOR TEST SECURITY: AN ARGUMENT FOR COMBINING PSYCHOMETRIC AND ONLINE PROCTORING METHODOLOGIES

Danielle Castelli, CompTIA | Russell Smith, Alpine Testing Solutions | Ashley Norris, ProctorU | Jaret Dyer, National College Testing Association
Panel Presentation

As technology evolves and progresses, test administrators have to be increasingly diligent in how they protect the security and integrity of their exams. Join thought leaders and experts on test security as they discuss methodologies and best practices to defend against proxy testing, cheating, and exam content theft. This session will provide multiple perspectives from test administration, test development and online proctoring. The panel will outline what they consider to be the strongest model for test security, technologies being implemented, psychometrics, supporting analytics, and speculation about the future of test security. In order to truly secure an assessment, mechanisms must be put in place to authorize test-taker identity, detect and prevent cheating in real time, and discriminate against content theft. Without employing a trusted method for all three of those concerns, an exam cannot be deemed completely secure. However, to date, there is no single method to defend against all three security concerns. This session brings together a panel of thought leaders and experts in test security, each with unique perspectives on test development methodologies, psychometric methodologies, online proctoring methodologies and how these methods can work together to provide the highest level of exam security. Panel participants will share best practices discovered and pitfalls to avoid based on their own experiences. They will bring data analytics to help define their experiences for the audience.

The session will follow a moderated discussion format with plenty of time for audience interaction. Whether you are new to the topic of test security or a seasoned expert in your field, everyone will learn something from hearing how these industry experts have worked to combine multiple technologies and methodologies from various professions to create the strongest model for test security available.

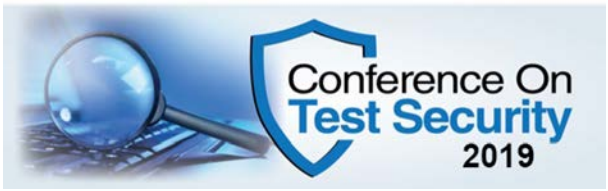
SALON C

EFFECTIVE SOLUTIONS FOR INVESTIGATING TEST SECURITY CONCERNS IN K-12 SCHOOLS

Chanon Bell, District of Columbia Office of the Statewide Superintendent of Education | Marc J. Weinstein, Caveon Test Security
Panel Presentation

State education agencies ("SEAs") face significant challenges when they identify test security concerns or potential irregularities in state assessment programs that require investigations to be conducted at the school level. Yet despite these challenges, SEAs must develop and employ effective solutions that provide for thorough and fair test investigations that enable the collection of trustworthy evidence upon which they can confidently determine what caused testing irregularities and whether anyone involved in the administration of the state assessment was negligent or engaged in potential test fraud. In this session, the presenters will describe solutions for school-level investigations that have been effectively used by state education agencies, including in instances where the state requires local education agencies ("LEAs") to conduct school-level investigations. Conference attendees who attend this session will learn how to: (1) decide which concerns and potential irregularities merit school- level investigations; (2) develop and implement specific policies and guidance that prescribe who must conduct investigations, and under what circumstances; (3) provide clearly defined standards and requirements for investigations; and (4) provide SEA and LEA staff with effective resources, guidance, tools and training to conduct investigations.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.



9:45 - 10:45 *CONTINUED...* *SALON D*

DATA FORENSICS: I WANT IT, I NEED IT, HOW DO I DO IT?

Craig Walker, Oklahoma State Department of Education | Michael Martin, Mississippi Department of Education | Tamara Lewis, Maryland Department of Education | Walt Drane, Caveon Test Security | Steve Addicott, Caveon Test Security
Panel Presentation

With test security scandals regularly splashed across newspapers, most state education agencies (SEAs) have implemented test security protections that help prevent, deter, detect and respond to test fraud (cheating and item theft). These protections help ensure the validity and reliability of their state assessment results. To this end, data forensic (DF) analyses are a powerful detection tool used by many SEAs to not only flag individuals, classrooms, and schools with high levels of unusual test taking that may involve test security violations, but also meet stringent federal Peer Review requirements. DFs highlights where and when unusual activity appears to have occurred, by whom, and the effects on test items. In this session, a panel of state assessment veterans showcase their use of data forensics that they have implemented into their assessment programs. Important topics to be explored include:

- What comprises your data forensics program? How was its implementation decided upon?
- How did policies and procedures need to evolve to support your data forensics program?
- How are data forensic results utilized, including score invalidation, informing investigations, highlighting training opportunities, and targeting test administration monitoring?
- How can I implement data forensics consistently across multiple test programs with multiple vendors?
- How can I defend data forensic results to parents, districts, or our Board?

The hard-won lessons shared by our panel will inform the audience, allowing them to leverage important test security protections for their own test programs while avoiding missteps along the way. A Question and Answer session will be available at the end of the presentation. Audience participation is highly encouraged. The goal for this session is for all participants to learn practical applications for using data forensics in their own state assessment program.

HIBISCUS

PREPARING YOUR PROCTORS

Ray Nicosia, ETS | Camille Thompson, The College Board | Mike Clifton, ACT | Rachel R. Watkins Schoenig, Cornerstone Strategies
Panel Presentation

Test administration is one of the most vulnerable phases of the testing process, with risks ranging from examinee cheating to test theft and item harvesting. Proctors stand on the front line in protecting your program's testing assets during the test administration phase. Given the decentralized nature of test administration, however, it can be difficult to ensure consistent proctor quality and attentiveness. How can proctors better prepare to deter or detect test security incidents? How can testing programs guide proctors on how to respond to concerns? Join testing professionals as they discuss emerging and time-tested methods to enhance proctor training and improve security for their testing programs.

Presenters will take a Demming-like approach and the premise that people want to do a good job proctoring. We will discuss ways to set forth expectations, including use of due diligence, contracting, and proctor manuals. We will also discuss how checklists, communications, and other messaging can help improve performance. Finally, we will discuss hotlines, targeted communications and standard reporting forms as additional tools to assist proctors. Participants will come away with practical and cost-effective ways to enhance proctor training and improve the "front line" of defense in protecting their testing assets.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

Conference On Test Security 2019

10:45 - 11:00

BREAK

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

11:00 - 12:00

SALON AB

STANDARD PRESENTATIONS I

Getting to Grips with Exam Fraud: A Qualitative Study Towards Developing an Evidence Based Educational Data Forensics Protocol

Sebastiaan de Klerk, Xquiry

This design research was focused on developing standards covering the entire process of examination to limit the chances of security risks (e.g., the prevention of exam fraud as much as possible, and detection by means of data forensics), together these standards form the Educational Data Forensics Protocol. Two research questions guided this study. The first question was, which standards regarding preventing and detecting fraud in the process of examination need to be included into the EDF protocol? In addition, practitioners must be able to act on indications of exam fraud based on these standards. Therefore, a second research question was formulated, namely which conditions must be considered during development of the EDF protocol to support practitioners in detecting possible gaps in the security of their examination process? The EDF protocol was developed and validated in five consecutive steps. This study analyses on the theoretical base of developing the EDF protocol (Step 1) and the considerations for developing a prototype (Step 2). The prototype was being validated (e.g., establishing correctness of the content) through seven semi-structured interviews with content experts in the field of either test security or data forensics (Step 3). Statements from these interviews were used to adjust the prototype into a final version of the EDF protocol (Step 4). Finally, to determine the practical value, the final version of the EDF protocol was used to flag gaps in the security of the exam process and determine possible security risks for one of eX:plain's exam programs (Step 5).

Assessment Data Forensics and Statistical Reporting and Consulting: Examples from Two Dutch Certification Programs

Erwin van Schaffelaar, Xquiry

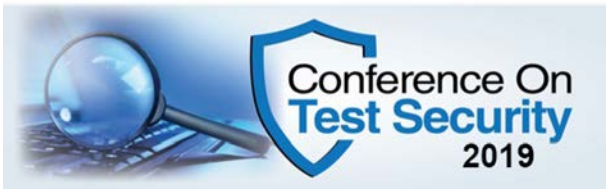
Xquiry offers assessment data forensics services for certification programs and educational institutions. A key element is the use of our assessment data forensics monitor. An intelligent, web-based application that can be used to detect aberrant patterns in test takers' response data. In the current presentation, we would like to share the experience that we have been building in the (statistical) reporting of the results of our assessment data forensics analyses for two Dutch certification programs. A corporate certification program for petrochemical and construction professions, and an educational certification program for Dutch special enforcement officers. We discuss the way to report findings from assessment data forensics analyses, as this is important to be effective as a statistical consulting company. The presentation will provide several examples from our reports, including the visualizations and tables we use to present statistical evidence. Besides the written reports, during the presentation we will also focus on two other elements of statistical reporting. First, how to get the (statistical) assessment data forensics message across in an explanatory conversation with the client. We discuss what works and what doesn't. Second, how to work together with your client to improve your statistical reporting. An important part in this process is to build a relationship based on trust with your client. Examples from our own experience will be given throughout the presentation.

Cut the Purse Strings: Using a Notice of Infringement to Credit Card Companies and Payment Processors as an Enforcement Tool

Brian C. Roche & Gerald C. Pia, Jr., Roche Pia LLC | R. Brent Hill & Brent Morris, Cisco Systems, Inc.

Although the internet provides test publishers with significant opportunity to promote, sell, and/or offer their exams globally, it also creates an opportunity for wrongdoers to pirate and sell those exams (or test-taking services) online. Respectable and legitimate payment processors and credit card companies are committed to preventing the use of their payment systems for unlawful transactions. Most will cooperate with IP Owners to investigate the infringing and other unlawful conduct of merchants utilizing their payment systems. In those instances, the rights of the IP Owners typically run parallel with the payment processors/credit card companies, who often view such investigation as critical to maintaining the trust of all the participants in their payment systems. The presenters, from Cisco Systems, Inc., and the law firm of Roche Pia LLC (a boutique cyberlaw firm based in Connecticut), together have a track record of coordinating with payment processors and credit card companies to combat IP infringement on the internet. The presenters will explain how to report IP infringing transactions to Visa, MasterCard, American Express, PayPal and other payment processors, and how those reports serve as a powerful IP enforcement tool by cutting off the pirate's ability to easily collect proceeds from the sales of their infringing products in U.S. commerce.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.



11:00 - 12:00 *CONTINUED...* *SALON C*

WHO'S MINDING THE CLOUD?

Benjamin Hunter, Caveon Test Security | Kirk Diepenbrock, American Board of Obstetrics and Gynecology | Rebecca Moden, Hewlett Packard Enterprise | Janet Lehr, Hewlett Packard Enterprise

Panel Presentation

It used to be that testing programs relied on in-house information technology staff to manage their sensitive data such as test items, item statistics, candidate records, and credentialing decisions. This data was supported and accessed only trusted personnel, using passwords, located in fire-proof, cardkey-accessed server rooms. The data was backed up and stored on a regular basis and sent to an offsite facility using tape media. In many of today's sensitive data management practices, testing programs have opted to move to computing Cloud environments, where their data is managed securely by third parties, offsite. We trust these outsourced entities to protect and manage our data securely. However, how do we know if our sensitive test data is really secure? What processes are in place to verify computing uptime? Who has access rights? And, who manages threats from bad actors with malicious intent? Join us as our panel of testing program managers and computing compliance expert discuss ways to manage testing program sensitive data in the Cloud. Topics discussed will include: how to work with third-party providers, understanding what safeguards and assurances should be put in place, and actions to take if a breach occurs.

SALON D

LATEST LESSONS LEARNED IN IMPROVING TEST SECURITY FOR STATE ASSESSMENTS: BEST PRACTICES AND RECOMMENDATIONS FOR THE PREVENTION OF CHEATING

John Olson, Caveon Test Security | Jessica Fenby, Michigan Department of Education | David Ragsdale, Massachusetts Department of Elementary and Secondary Education | John Fremer, Caveon Test Security | Dusty Shockley, Delaware Department of Education

Panel Presentation

Over the past five years, many lessons have been learned on improving test security for state assessment programs. States have become much more proactive in their approaches for implementing stronger procedures to prevent cheating. As the old saying goes, an ounce of prevention is worth a pound of cure, and many State Assessment Directors have been dishing out pounds of preventive medicine in recent years for more secure testing programs. In this session, information will be provided from a wide variety of perspectives. Presenters will discuss how things have improved since two important documents were written and shared with the state assessment community. These are the *TILSA Test Security Guidebook for States* and the *Lessons Learned in Improving Test Security for States*. These two reports garnered much attention from states and vendors by providing numerous examples of approaches that have been implemented to stop cheating on tests and bringing together many of the best practices and procedures of state staff and vendors. Because 4-6 years have passed since these reports were published and much has happened since then, recent lessons that have been learned by states on test security will be shared. Presenters will describe current and planned activities on several important themes related to test security, (a) recommended methods/approaches/guidance to improve test security in states, (b) meeting federal requirements for peer review, especially for test security and monitoring of assessment administrations, and (c) the latest lessons learned and best practices for preventing cheating.

HIBISCUS

NCME'S POSITION STATEMENT ON TEST SECURITY: A GOOD START, WITH CONSIDERATION OF CLARIFICATIONS AND AREAS FOR IMPROVEMENT

Steve Ferrara, Cognia | David Foster, Caveon Test Security

Panel Presentation

In March of this year, the Board of Directors of the National Council on Measurement in Education (NCME) approved a Position Statement on Test Security. The statement asserts that "rigorous and effective policies and practices are necessary to protect the security of test content and the integrity of test scores. Test security violations undermine the validity of intended score interpretations and uses, which can lead to faulty decisions based on test scores."

This panel presentation features two active advocates and practitioners for test security policies and practices. They share decades of experience in preventing, detecting, investigating, and resolving test security violations for high stakes, operational testing programs. The panelists will discuss the position statements - both its strengths and areas for improvement. The goals of this panel discussion are to evaluate the position statement, identify strengths weaknesses, recommend needed improvements, and expose the statement to the audience of experts at the Conference on Test Security, all with the goal of improving the statements. The panelists will actively promote audience participation in the discussion.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

12:00 - 1:00

SALON F

LUNCH

1:00 - 2:30

SALON AB

BAKING IN TEST SECURITY FROM THE START: A STRATEGIC RECIPE FOR ENHANCING EXAM PROTECTION

Benjamin Hunter & Saundra Foderick, Caveon Test Security
Panel Presentation

Creating secure and fair exams is much like baking the perfect cake: you need to use quality ingredients and follow the steps in the right order or you will have less than desirable results. This session will describe proactive test security strategies and practices that can help testing programs create secure quality assessments, yielding test results that can be trusted. We will discuss ways to limit item exposure, use innovative item types to reduce risk and deter theft, increase score validity, measure test security effectiveness, and possibly track those parties responsible for exam exposures or theft. Examples will showcase easy-to-implement techniques, including digital watermarking, innovative item types, secure exam design check off sheets, item writing strategies, and post-delivery tracking and analysis. The recipe for protective test security does not start when you deliver an exam or end when an exam is published; protective test security practices follow a strategic recipe that is baked into your testing process from the beginning.

HIBISCUS

DO YOU HAVE AN EFFECTIVE COMMUNICATIONS PLAN FOR HANDLING A MAJOR TEST SECURITY BREACH?

Camille Thompson, The College Board | Alison Foster, Caveon Test Security | Ray Nicosia, ETS | John Fremer, Caveon Test Security
Panel Presentation

Major test security breaches can occur at any point in the life cycle of an exam or item pool. Often, the event is discovered during a test administration. Subsequent questions can come from test users, test administrators, and the media. Many entities want and need answers—good answers; ones that will stand up to close scrutiny. What is available in your testing program at this time? At best, you may have some or all of the following: i) a Security Incident Response Plan, ii) a SIRP Coordinator, iii) a backup test form or item pool, ready to be deployed, iv) a communication team ready to collect information and coordinate communications, using a single point of contact, v) many other important components of a Security Incident Response Plan.

In this session, experts from ETS, The College Board, and Caveon Test Security will provide valuable information to assessment programs in successfully managing and communicating before, during, and after a security breach. Among the areas that will be addressed are: i) policies and procedures, ii) training, iii) legal issues, iv) the different perspectives to be considered, v) common pitfalls, vi) lessons from past incidents.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

1:00 - 2:30 *CONTINUED...* *SALON C*

STANDARD PRESENTATIONS II

A Picture Is Worth A Thousand Words: Using Data Visualization to Find and Communicate Patterns of Test Fraud

Nicole Tucker, Zuru Du, & Greg Hurtz, PSI Services

A successful test security program includes procedures for clearly communicating the nature and extent of a potential breach or cheating incident. Methods for detecting potential test fraud often rely on statistical indices that may be difficult for decision makers to interpret. Data visualization provides an opportunity to graphically tell the story clearly and efficiently, in a way that non-statisticians can often clearly see the anomaly that our statistical indices detect. This presentation will review a proven suite of data forensics indices along with accessible yet informative graphs and plots. These graphs and plots have been designed with care in order to visually represent data trends, to detect and communicate patterns consistent with test fraud. Specifically, pass rate elevations, candidate similarity (J2), modified caution index (MCI), speed, and changes in item performance are used as flagged indicators against a baseline of expected performance, and often plotted against test scores to help visualize their potential impact on testing outcomes. Examples will be presented of graphical representations of examination results consistent with normal testing behavior as well as confirmed incidences of test fraud. Knowing what to expect from normal behavior, and then having prototypical examples from confirmed cases, helps recognize the emergence of potential anomalies that can be used for early detection.

Call and Response: Test Security Process, Analysis, and Client Communication Following Triggering Events

Brooke Dresden, Bryan Byington, Nicole Tucker, & Greg Hurtz, PSI Services

An exam proctor reports observation of rapid test-taking among state licensure candidates. A quarterly exam statistics report reveals apparent abnormalities in a prelicensing school's pass rates. This presentation will follow along with these two separate state licensure examinations as they experience triggering events for a test security investigation. We will outline the procedures in place to identify potential triggering events, and our response to those events, including analysis and client follow-up. We will provide a high-level overview of the group- and candidate-based analyses performed in response to the triggering events and discuss how to present the results to the client in a way that is informative, succinct, and accessible.

Lessons Learned: Leveraging Technology to Efficiently Evaluate Potentially-Compromised Test Items on Tests Containing Foundational and Clinical Science Content

Kimberly Swygert, Ian Micir, & Linda Adler, National Board of Medical Examiners

The interpretations of test scores in secure, high-stakes environments are dependent on several assumptions, including that examinees are not exposed to items in advance of the test. Item harvesting and sharing threaten programs that are high-stakes, contain relatively short and memorable multiple-choice questions (MCQs), or that have continuous testing windows with item re-use. This presentation explores the development and implementation of a new process intended to evaluate the presence and impact of suspected item harvesting/sharing on multiple exams related to foundational and clinical science content. The following will be covered: (1) defining what it means for scored test materials to be impacted by item harvesting/sharing, (2) establishing efficient triaging methods for matching harvested material to scored material, (3) development of a specialized software application that uses natural language processing and cosine similarity indices to support a fast, accurate comparison of large segments of potentially compromised material to scored material, and (4) defining a response plan to evaluate and mitigate the impact of any identified compromised items. The primary focus of the presentation will be on step (3), software development, as this is most innovative part that is unique to our organization. Our goal with this presentation is to educate the audience on the lessons learned during the development of the software and the processes surrounding it, in an effort to inform discussion among industry professionals about the critical considerations for evaluating the presence of item harvesting/sharing for any exam program and how they can take steps to mitigate the impact.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

1:00 - 2:30 *CONTINUED...* *SALON D*

STANDARD PRESENTATIONS III

An Unsupervised-Learning-Based Approach for Detecting Compromised Items

Yiqin Pan & James Wollack, University of Wisconsin

As technologies have been improved, item preknowledge has become a common concern in the test security area. The present study proposes an unsupervised-learning-based approach to detect compromised items. The unsupervised-learning-based compromised item detecting approach contains three steps: (1) classify responses of each examinee as either normal or aberrant based on both the item response and the response time; (2) use a recursive algorithm to cluster examinees into groups based on their response similarity; (3) identify the group with preknowledge and report as compromised those items to which most examinees in this group give aberrant responses. Results show that under the conditions studied, provided the amount of preknowledge is not overwhelming, the approach controls the false negative rate at a relatively low level and the false positive rate at an extremely low level.

Detecting Item Preknowledge in Real, Marked Data via Graph Theory Approach

Dmitry I. Belov, Law School Admission Council | Sarah L. Toton, Caveon Test Security

An experiment was conducted to embed item preknowledge into a small group of examinees. Examinees in the control condition (33) simply took a 25 item multiple choice test, but examinees in one experimental condition (30) had access to a subset of 12 items with answers before the exam. The resulting real dataset was used to study properties of a recently developed graph theory approach to detect examinees involved in item preknowledge. The response similarity index Omega by Wollack (1997) was used to construct a graph that represents the structure of similarity among examinees and then characteristics of this graph were assessed. The results showed that the graph is sensitive to item preknowledge and that using an uncompromised subset of items to estimate ability dramatically increased detection rates.

Validation of Graph Theory Approach to Detect Test Collusion Network Using an Experimental Real Dataset

Cengiz Zopluoglu, University of Miami | Dmitry Belov, Law School Admission Council | James A. Wollack, University of Wisconsin

A graph theory approach has been introduced and applied to detect group of test takers who are collectively involved in test collusion (Belov and Wollack, 2018). In this study, we explore the utility of this new approach using a real experimental dataset. Participants were 163 students selected from 40 classrooms of 16 schools in a medium metropolitan area of southern Greece. The experimental manipulation took place in students' regular classrooms from their respective teachers. The teachers administered a mathematics test at the beginning of a two hour lesson and informed students that a second part of the initial test would be administered at the end of the lesson, in order to first cover some material. The first part of the test was administered under close and strict invigilation. Half an hour before the end of the lesson, the teacher administered the second part of the test. During the second test which was identical in difficulty, complexity, and number of exercises with the first exam, the invigilator pretended that he/she received a very important call and had to run to the principal's office. He/she further told students to continue with the test while he/she would leave the room for a few minutes. At last, the teacher returned to the class and collected the second part of the test. The results revealed some interesting networks of students who potentially collaborated during the experimental condition while there was not any clique identified in the control group. We provide a closer look to the similarity between the response patterns among the individuals in identified cliques

A Method for Real-Time Anomalous Response Detection for Computer-Based Linear Tests

Merve Sarac & James Wollack, University of Wisconsin

We have approached forensics analysis in a way that we wait until the end of a test administration, run forensics, investigate any potential misconduct, and then make decisions, such as canceling scores and requiring a retest. However, it is time-consuming and expensive, and, in cases where suspicious examinees need to schedule a retest, delays reporting of valid scores. While there is considerable research supporting the use of methods to detect anomalous responses applied after a test is completed, no studies exist to support their use in real-time. Real-time anomaly detection offers the opportunity to make decisions during a test so we can either terminate the exam or route the candidate towards different content that might yield a credible test score. This study describes and evaluates a method for real-time anomalous response detection using a score differencing index within the context of item preknowledge. An extensive simulation study is conducted in which proportions of compromised items, proportions of examinees with preknowledge, levels of anomaly detection periodicity, starting-points at which to begin anomaly detection, and ending-points at which to stop anomaly detection were all manipulated. Type I error rate and power are computed to evaluate the results of each condition. We also report differences across conditions in how early examinees were flagged.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

Conference On Test Security 2019

2:30 - 3:00

EXTENDED BREAK

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

3:00 - 4:00

SALON AB

EXAM SECURITY CLIFF NOTES

David Ragsdale, Massachusetts Department of Elementary and Secondary Education | Jim Sconing & Rachel R. Watkins Schoenig, Cornerstone Strategies | Joe Grochwalski, The College Board
Panel Presentation

Z-scores. P-values. Standard deviations. Surrogate testing. Item harvesting. For most of the population, these terms sound like you're practically speaking a foreign language. To be effective, exam security experts must cross the language divide to communicate with a wide range of non-experts, from executives to attorneys and judges to the media. How can we help decision-makers better understand the value and weight of statistical evidence and the issues we are trying to address? What types of visuals can be presented to examinees, attorneys, arbitrators, and judges to better explain the evidence and exam security concerns? Join seasoned exam security experts as they use plain language and case examples to present methods that effectively move from the abstract and general to the specific and concrete when dealing with exam security concerns. If you ever have to explain exam security issues to your executives, examinees, or others, this is one session you won't want to miss.

SALON C

STANDARD PRESENTATIONS IV

Comparing B3 to Answer Similarity Index for Detecting Collusion

Russell Smith, Alpine Testing Solutions

Foley (2019) proposed using a variation of Yen's Q3 statistic (1984) for collusion detection, calling it B3. Where Yen's Q3 statistics uses item residual correlations to detect dependencies between pairs of items, the B3 statistic uses person residual correlations to detect dependencies between pairs of test takers. This paper will compare and contrast the advantages and disadvantages of the B3 statistic to Answer Similarity Index (ASI) analyses using the generalized binomial model (van der Linden & Sotaridona, 2006). The B3 statistic has a few potential advantages for detecting collusion. It is readily available via commercial software (e.g. Winsteps, Linacre (2019)), it is generalizable to polytomous models, and it accounts for in both the number of matching scores and the likelihood of the scores given the test takers' abilities and the difficulty of the individual items. One potential disadvantage of the B3 statistic is that the underlying distributions for B3 is not known making it challenging to estimate probabilities. Real data with known exposure problems as well as simulated data will be used to evaluate the proposed statistic and its viability by specifically comparing it to ASI.

Collusion Detection Using an Extension of Yen's Q3 Statistic

Brett P. Foley, Alpine Testing Solutions

Collusion can be indicated when two or more examinees perform more similarly than one would expect after accounting for their ability and the difficulty of the items they answered. In this presentation we introduce a new metric that holds promise for identifying this type of cheating. The metric is based on Yen's Q3 statistic, can be calculated for tests comprised of dichotomous and/or polytomous items, and can be estimated by widely-available IRT software (e.g., WINSTEPS). Attendees will be shown how the statistic is calculated and provided with explanations of the method's strengths and weaknesses. Results from preliminary applications of the method will be shown.

The Prevention and Detection of Test Fraud: Two Empirical Studies on Assessment Data Forensics

Sebastiaan de Klerk, Xquiry

In this session, we present the results of two research studies on the prevention and detection of test fraud. The first study is a quantitative empirical study on the reliability of two assessment data forensics parameters: the Guttman error person-fit parameter (Meijer, 1994), and a lognormal response time model parameter (Van der Linden, 2006). We analyzed and compared aberrant response patterns for these parameters for five experimental conditions of participants instructed to cheat on a test (i.e., collusion, proctor assistance, cheat sheet, smartphone use, and pre-knowledge) and a control group of honest test takers. Results show that combining the person-fit parameter and the response time parameter leads to the most optimal judgment. Although the detection ratio was still rather low - 38% of instructed cheaters were detected -, the reliability was high - the true positive ratio was 97%. This means that detected test takers are very likely to have cheated. The second study is a qualitative study on an assessment data forensics protocol. The protocol is a self-developed appraisal system consisting of evidence-based guidelines on the prevention and detection of test fraud. A prototype of the protocol was subject of eight semi-structured subject matter expert interviews. Furthermore, the protocol has been applied in a Dutch testing organization. Based on the interviews and the test case, the prototype has been adjusted into its final form. The results will be presented.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

3:00 - 4:00

CONTINUED... *SALON D*

Did You Know There Is An ITEMS For That?

Steve Ferrara, Cognia | Sarah Toton, Caveon Test Security | Frank Padellaro, Cognia | Cengiz Zopluoglu, University of Miami
Panel Presentation

ITEMS or Instructional Topics in Educational Measurement Series, is just that: self-paced, multi-media modules of instruction around various measurement content areas sponsored by the National Council on Measurement in Education (NCME). The benefit of these modules? They are engaging, well thought out lessons taught by experts in the field. The beauty of these modules? They are free to anyone who wants to learn more about educational measurement.

The latest ITEMS modules are being developed for the topic of Test Security, with such content areas as: comprehensive test security programs, understanding data forensics, theft-resistant item development strategies, and protecting test content by web and media monitoring. This session will discuss the various test security topics that will be covered in ITEMS and the education treatments that will be applied. The speakers will solicit feedback from participants to help finalize the ITEMS pedagogy.

HIBISCUS

A STATEWIDE SYSTEMATIC APPROACH TO TEST SECURITY

Sholonda Trice, Indiana Department of Education | Molly Chamberlin, Chamberlin/Dunn LLC | Wes Bruce, Independent Consultant | Andrew Bernlohr, Indiana Department of Education
Panel Presentation

Indiana approaches test security both proactively and from an incident response perspective. Indiana formalized test security protocols based on requirements from federal Peer Review which require monitoring of state testing sites throughout the standardized assessment window. Additionally, response analyses are undertaken following the test administration window which highlight aberrant administration practices requiring additional investigation. Four presenters will highlight different aspects of the monitoring and test security protocols undertaken by Indiana. Specifically, the presentation will note the historical landscape associated with test irregularities and test security investigations in Indiana, recent changes based on procurement of additional systems which now offer new data forensic analysis points, and common areas of deficiencies and lessons learned as states consider implementation of a large scale monitoring process for statewide assessments. These analyses peel back successive layers of data to reveal patterns of student response that inform the concerns raised by the statistical analytics. We seek to highlight implications other states may consider as they engage in these types of processes to ensure testing integrity is maintained.

4:00 - 4:15

BREAK

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

4:15 - 5:15

SALON AB

REMOTE, ONLINE PROCTORING: TRENDS, BEST PRACTICES, AND PRACTICAL APPLICATIONS INTO THE FUTURE

Benjamin Hunter, Caveon Test Security | Steven Winicki, Examity
Panel Presentation

Innovation fuels great change in every industry, and the testing industry is no different. Technology has changed the way that we've assessed knowledge since we began assessing, in search of competitive advantage, lower cost, and better outcomes. The act of proctoring has undergone a similar, parallel evolution and will continue to evolve as market pressures exert themselves in new and exciting ways. This session will:

i. Explore the evolution of remote, online proctoring
ii. Compare and contrast best practices in proctoring, identifying the ways that different proctoring methods share commonality
iii. Look at trends in online proctoring as it moves from humble beginnings in the higher education and distance learning markets, towards the certification and licensure markets
iv. Identify and evaluate hot-button legal and privacy issues that are rapidly evolving as legislators try to keep up
v. Discuss technological advances in use today that support the adoption of online proctoring, and muse about (currently) theoretical innovations and how the development of those innovations could have practical application in the future

Salon C

DELVE INTO THE DIGITAL DOMAIN: MONITORING YOUR PROGRAM'S E-PRESENCE

Amy S. Ressler, The College Board | C. Beth Hill, National Conference of Bar Examiners | Jen Baldwin, Caveon Test Security

Our world has become increasingly connected and digital in nearly every facet of modern life. From socializing to shopping, dating to doctor's visits, our daily activities are being incorporated into the online arena. The testing industry is no different. Testing programs now build tests using online software, administer tests with on-demand testing, and employ online proctoring services, etc. However, the internet is an unruly and ever-evolving environment, and while it brings many advantages, it has also opened new opportunities for people to expose and discuss test content online. The purpose of web monitoring is to survey this ever-changing virtual landscape, to understand online trends that affect the testing industry, and to proactively tackle internet hotspots for sharing and exposing exam content. One of the most effective and efficient ways you can protect your organization is to understand how these digital trends affect your program's specific online presence, and then develop a robust and fluid web monitoring strategy to address them.

SALON D

STANDARD PRESENTATIONS V

A Deterministic Gated Lognormal Response Time Model to Identify Examinees with Item Preknowledge

Murat Kasli & Cengiz Zopluoglu, University of Miami

There is a growing interest in the literature about response time information as it may provide additional understanding about item preknowledge. In this study, a Deterministic Gated Lognormal Response Time (DG-LNRT) was proposed with the purpose of differentiating aberrant response behavior from the normal behavior. A simulation study will be presented to evaluate the performance of the model in identifying examinees with item preknowledge. In the simulation study, four different variables will be manipulated: sample size, number of items, percentage of compromised items, and the percentage of examinees with item preknowledge. Additionally, potential expansion of the model for van der Linden's hierarchical IRT model using both raw item responses and response time information will be discussed.

4:15 - 5:15 *CONTINUED...* *SALON D*

Detection of Item Preknowledge by Exploiting Uncompromised Subset of Items

Dmitry I. Belov, Law School Admission Council

General case of item preknowledge (IP) is studied, where groups of examinees (aberrant examinees) had access to subsets of items (compromised items) from an administered test prior to the exam. Nothing is known about these groups and subsets. When only item scores are given then, according to Karabatsos (2003), the highest detection of aberrant examinees is achieved with TW statistic developed by Trabin and Weiss (1983). This proposal exploits a common situation in high stakes exams when uncompromised subset of items is known (e.g., items administered for the first time) and modifies TW to take advantage of this subset. Then this proposal develops a new Monte Carlo approach for detecting aberrant examinees by estimating mean of a random variable measuring a score gain from uncompromised subset to a specially constructed subset from the test such that a nonaberrant (an aberrant) examinee should have a low (a high) score gain. Comparison study with TW and other IP detectors using data simulating the general case of IP demonstrates advantages of applying the Monte Carlo approach. Further applications of the approach are discussed.

Using the Likelihood Ratio Test Statistic in Detecting Item Preknowledge

Aranka Bijl, Xquiry

The testing community is moving towards Computer-Based Testing (CBT) at an increasing pace. However, using the necessary item banks in CBT, a challenge arises that items get increasingly exposed to small groups of test takers. This increases the likelihood of an item being compromised. Item preknowledge is a form of cheating, defined as a situation where a group of test takers have had access to one or more items and/or their answers prior to the test administration. The presumption exists that test takers with item preknowledge show different behavioral patterns as compared to honest test takers. This is apparent in extremely short response times on correctly answered items and correct responses on items outside of the test takers ability. For the current empirical and experimental research, we investigated whether the Likelihood Ratio Test (LRT) statistic (Sinharay, 2017) is able to detect a known sample of test takers who had preknowledge to a subset of items in a test.

HIBISCUS

RECENT ACTIVITIES TAKEN BY STATES PER FEDERAL PEER REVIEW REQUIREMENTS ON TEST SECURITY AND TEST ADMINISTRATION MONITORING

John Olson, Caveon Test Security | Chris Seay, South Carolina Department of Education | Craig Walker, Oklahoma Department of Education | Timothy Butcher, West Virginia Department of Education | John Fremer, Caveon Test Security

Coordinated Symposium

In 2018-2019, the USED provided feedback to states on how well they met the new peer review requirements for test security, and in particular, their monitoring of test administrations. Over the past few years all states were required to submit evidence to the feds that their assessment systems have integrity and that the tests are administered in a secure manner. Not all states met these requirements. Many received feedback from USED that identified areas where they were either lacking in evidence or in documenting the procedures being used. Peer Review Critical Elements 2.4 and 5.4 on monitoring assessment administrations have been identified as problematic areas for many states. In its formal feedback to states in early 2019, the USED stated that peer reviewers found many states had not provided adequate evidence that they have "implemented with fidelity and documented an appropriate set of policies and procedures to prevent test irregularities and ensure the integrity of test results."

In this session, three states will describe the types of evidence they submitted for peer review that address the test security requirements. They will also describe how they used the feedback they got from USED on their submissions for making further improvements to their approaches. In addition, a nationally-recognized assessment expert and an experienced peer reviewer will discuss the different types of evidence from states that has been found to be particularly supportive of secure state assessment systems and recommend various ways that states can make additional improvements to their programs.

THURSDAY, OCT 17
7:45 A.M. – 8:00 P.M.

6:00-8:00

POOL DECK LOWER (5th Floor)

COCKTAILS & CONVERSATIONS

POSTER EVENT

Light hors d 'oeuvres will be served

You won't want to miss it!

POSTER PRESENTATIONS

Using New Items to Detect Cheating Behaviors in Computerized Multistage Testing
Yibo Wang, Deborah Harris, & Stephen Dunbar, University of Iowa

The Effects of Multiple Comparison Adjustment Methods on Detection of Answer Copying
Hui Deng, College Board

Detection of Test Collusion Among Multiple Examinees
Hongling Wang & Chi-Yu Huang, ACT

On Group Invariance of the Empirical Null Distribution for Similarity Analysis
Zebing Wu, City University of New York | Zhongmin Cui, ACT

On a New Data Cleaning Method for Similarity Analysis
Mingjia Ma, University of Iowa | Zhongmin Cui, ACT

Conditioning on Raw Scores in Investigations of Answer Similarity
Carol Eckerly, ETS

On Creating an Empirical Null Distribution for Similarity Analysis in the Absence of Large Data
Zhongmin Cui, ACT

Using Student Videos to Ensure Test Security
Walt Drane, Caveon Test Security | Dusty Shockley, Delaware Department of Education

Parents' Willingness to Help their Child Cheat on College Admissions Exams
John Jones & Kelly Dages, FifthTheory

David vs Goliath (small certification programs vs large certification programs)
Kelli Foster & Susan Weaver, Caveon Test Security

FRIDAY, OCT 18
8:00 A.M. – 1:00 P.M.

7:00 - 8:00

SALON F

CONTINENTAL BREAKFAST

8:00 - 9:00

SALON AB

DATA FORENSICS OVER TIME: ANALYSIS TECHNIQUES AND METHODS

Coordinated Symposium

Using Relative Score Changes to Detect Potential Test Fraud

Sarah L. Toton & Dennis Maynes, Caveon Test Security

Score changes over time are an important indicator of student learning, but dramatic score changes are unusual and may indicate that test fraud has occurred. Often score changes are assessed by predicting the most recent score from the earlier score using ANCOVA, RMANOVA, or regression. However, using these methods, high performing schools are consistently flagged over time, likely due to regression to the mean among other schools. In this presentation, we propose using a latent trait model to analyze score changes, which addresses this issue and is appropriate for detecting potential test fraud. We will compare results using a regression-based method, predicting recent scores from an earlier score, and the latent trait method in a real dataset that has been manipulated.

Evidence of the Proliferation of Braindump Content in Real Data

Marcus W. Scott & Dennis Maynes, Caveon Testing Security

Delivering tests over the Internet has created a situation where test thieves can steal exam questions, but not their answer keys because the answer keys are not downloaded to the test site server. Experience has shown that test thieves who then create their own answer key often make errors, resulting in a flawed answer key for the braindump. Examinees who frequently select the incorrect answer in the flawed key likely used the braindump. Several methods for detecting examinees who used a known flawed answer key exist. When braindump users can be reliably detected, then proliferation of the braindump through a testing population can be tracked. Additionally, the braindump's effects on item p-values, test scores, and pass rates can be quantified. This presentation presents results on braindump proliferation and its effects in real-life data sets.

Bayesian Methods for Score Differencing

Sandip Sinharay, ETS

Wollack and Schoenig (2018) categorized the statistical methods to detect test fraud into six categories, one of which is "score differencing". This category of methods involve a test of the hypothesis of equal ability of an examinee over two sets of items against a one-sided alternative hypothesis. Score differencing can be performed to detect several types of test fraud including fraudulent and large gain scores where the two sets of items were administered at two different time points. Bayesian methods have rarely been applied in score differencing, with the exception of Wang, Liu, and Hambleton (2017). This paper suggests two new approaches to perform score differencing using Bayesian methods. One approach is based on posterior probabilities and the other on Bayes factors. The false alarm rate and the power of the two approaches are examined using simulated data. Then the approaches are applied to a real data set.

Using Predictive Pass Rate to Monitor Potential Group Cheating

Yu Zhang, Yi Lu, Aijun Wang, & Lorin Mueller, The Federation of State Boards of Physical Therapy

Collusion is a common incident in test fraud and it can happen for test takers who act as a group or team up in different ways: by test center, school, test preparation course, or online. The test takers from the same group might recall or share items, and then gain advantage in their test performance. To detect group cheating is a critical task for test developers to ensure the validity of test scores. This study introduces a group-level analysis of potential cheating, group-level pass rate predictions for a licensure exam. The goal of this analysis is to use information about test takers to better understand, ultimately predict, and monitor aberrant fluctuations of group-level pass rate. Known predictors of passing probability include previous scores, previous attempts, and scores on the practice tests.

FRIDAY, OCT 18
8:00 A.M. – 1:00 P.M.

8:00 - 9:00

CONTINUED... *SALON C*

TEST SECURITY AND STUDENT EXPERIENCES PRESERVING THE INTEGRITY OF THE DEGREE WHILE SUPPORTING STUDENT SUCCESS

Lestelle Greenwalt & Carissa Pittsenberger, Western Governors University
Panel Presentation

This session explores the multi-pronged assessment security approach within the Western Governors University, highlighting two components: the Take a Break feature and Secret Shopping. The approach ensures assessment security as well as a positive student experience through the protection of the integrity of assessment results and the Western Governors University brand.

SALON D

YOUR EXAM DATA HAS BEEN BREACHED - WHAT DO YOU DO?

Bryan Friess, Pearson | Micheal Clifton, ACT | Brent Morris, Cisco Systems, Inc.
Panel Presentation

Your exam booklets are lost, your exam content is being shared online, you have reports of candidate data being exposed - what do you do?
These types of reports and incidents are indications that a problem has presented itself and could be precursors to a much more serious issue (Just ask any organization that has had a data breach in recent years). Having an incident response plan in place will prepare your organization to deal with these types of things when they occur. In this session industry experts will share different mechanisms to monitor and identify potential incidents, respond to an incident in a timely manner, analyze the situation and determine communication plans, take measures to protect data, prepare for business recovery post incident damage, and retrospect on the process.

HIBISCUS

EVIDENCE FOR THE TEST SECURITY BENEFITS OF SMARTITEMS

David Foster, Caveon Test Security | Arthur Altman, SailPoint
Panel Presentation

While providing other benefits, SmartItems were invented in early 2018 for the purpose of preventing item theft and most forms of cheating. After a year and a half, what evidence is there that they work as promised? Rationally, if they are used on tests, theft is impossible, along with insidious forms of cheating, but is that logic supported by research findings? Since they were introduced, there have been gathered three sources of evidence regarding their effectiveness: data simulations, case studies involving actual certification exams, and scientific experiments. Individually or combined, these sources indicate that SmartItems exceed psychometric performance standards, while simultaneously preventing the piracy of exam content, and almost every form of cheating. The results of these studies will be presented and discussed.

9:00 - 9:15

BREAK

FRIDAY, OCT 18

CLOSING KEYNOTE

9:15 – 10:30 AM | SALON E

EXAM SECURITY THROWDOWN

Moderator: Kim Brunnert, Elsevier

It's all fun and games until you learn something! Industry experts will team up to test their mettle in responding to questions across a range of test security related topics. Be part of the game-show action and enjoy an hour of laughter and learning with your colleagues and friends across the industry. Trust us – you won't want to miss this!

FRIDAY, OCT 18
8:00 A.M. – 1:00 P.M.

10:30 - 10:45

BREAK

10:45 - 12:00

SALON AB

STANDARD PRESENTATIONS VI

Impact of Cheating on IRT Equating of Licensure Exams

Juan Chen, Mengyao Zhang, & Mark A. Albanese, National Conference of Bar Examiners

A simulation study was conducted to investigate the impact of cheating on IRT true-score and observed-score equating based on data from a large-scale licensure exam. Various levels of cheating were simulated through manipulation of the proportion of compromised items, the percentage of cheaters, and types of cheating. Equating results and pass/fail decisions were obtained for the cheating conditions and were compared to those of a baseline condition that involved no cheating. It is expected that high levels of cheating could lead to artificially inflated equated scores.

A Method for Adjusting DOMC Scores to Account for Test Difficulty

Dennis Maynes, Caveon Test Security

Discrete Option Multiple Choice (DOMC) items have considerable test security advantages that prevent disclosure and use of the disclosed content. However, recently concerns have been expressed about the fairness of tests comprised of DOMC items because the items can vary in difficulty and discrimination, producing tests that vary in their psychometric characteristics. This research demonstrates that variability in the difficulties of DOMC items can be computed and accounted for in test scores when DOMC items are used, resulting in test scores that account for test difficulty and allow comparison across examinees. The mathematics of the method follow from sequential probability computations, which means the overall method is a straightforward application of standard statistics and probability theory. Furthermore, this research compares how well the raw, adjusted, and true scores align for both DOMC items and MC items.

The Use of Item Parameter Drift and Multi-Facets Rasch Model to Detect Case Exposure in a High-Stakes Objective Structured Clinical Examination (OSCE)

Karen Coetzee, Touchstone Institute

The epidemic of candidate cheating within high-stakes testing contexts is perhaps particularly worrisome within the medical licensure field where assessments serve predominantly as gatekeepers for public protection. The popular Objective Structured Clinical Examination or OSCE test format used within this field heightens the risk for candidate cheating via ease of information sharing prior to the examination. At the same time, tracking shifts in item parameters over time, a phenomenon commonly known as item parameter drift (IPD), to help detect potentially exposed test content is further complicated in OSCE settings, given the variance associated with the need for examiners to assign scores. This study specifically applied a four Facet Multi-facets Rasch model (MFRM) to investigate IPD patterns of 36 high-stakes OSCE cases administered over several administrations. Results revealed patterns of significant IPD in terms of increased case easiness in eight of the 36 stations (22%), indicating potential exposure. Retirement of these cases is recommended and the adjustment of candidate scores based on these results warrants further investigation. The generated difficulty estimates for the remaining 22 cases provides a convenient system for tracking and monitoring IPD patterns in future administrations.

Application of a New Method for Multiple-Group Analysis of Jointly Modeling the Item Responses, Response Times, and Visual Fixation Counts.

Kaiwen Man, University of Maryland | Jeffrey R Harring, University of Maryland | Cengiz Zopluoglu, University of Miami

Many approaches have been proposed to jointly analyze the item response and response times to understand the behavioral differences between the normally and aberrantly behaved test-takers. However, the biometric information collected from the innovative technology-enhanced learning system (ITELS) is left behind. Given this context, this study demonstrates the application of a new method for multiple-group analysis of jointly modeling the item responses, response times, and visual fixation counts collected from an eye-tracker. The behavioral differences between the normally behaved test-takers and the ones who have pre-knowledge about the test items will be manifested via the trade-offs among their latent ability, the working speed and the testing engagement by a person-side variance-covariance structure. Bayesian estimation scheme is used to fit the proposed model to data, and the results are discussed.

FRIDAY, OCT 18
8:00 A.M. – 1:00 P.M.

10:45 - 12:00

CONTINUED...

SALON C

HOW A NURSING CERTIFICATION PROGRAM USES PSYCHOMETRICS, DATA FORENSICS , AND WEB PATROL TOGETHER TO PROTECT THEIR PROGRAM

Tim Sares, American Nurses Credentialing Center | Sarah Toton, Caveon Test Security

High stakes testing programs go to great lengths to secure the content of their certification exams and ensure the validity of their candidates' scores. In the medical arena, this takes on added importance due to the necessity of protecting the public from fraudulently obtained certifications. Psychometrics, data forensics, and web patrol are three important components of a test security plan to protect the validity of test scores and their interpretations. These are typically viewed as entirely separate processes, but all three contribute to the overall health of a program and often provide overlapping sources of evidence in identifying or investigating a security breach. This session will cover the type of information provided by each of these areas as well as how integrating this information can lead to a stronger testing program. The presenters are a psychometrician who works in the data forensics area of a test security firm and a psychometrician who works for a nursing certification board, who is responsible for test security planning, investigations, and responding to security threats or incidents for several testing programs. They will speak about real cases in which psychometrics, data forensics, and web patrol have played crucial roles in identifying or investigating breaches and how they fit into a comprehensive test security plan.

SALON D

MAKING THE MOST OF MONITORING

Jessica Fenby, Michigan Department of Education | Jeff Holtz, Minnesota Department of Education
Panel Presentation

Making the Most of Monitoring is a two state panel discussion about K-12 assessment administration monitoring practices. Representatives from Michigan and Minnesota will discuss the planning, implementation, and review practices used to make the most out of K-12 assessment administration monitoring. Panelists will discuss current assessment monitoring practices used in their respective states for test security.

HIBISCUS

GOING BEYOND INTERNAL TEST SECURITY RESOURCES: REACHING OUT FOR EXTERNAL ASSISTANCE

Cathy Koenig, American Board of Pediatrics | Mark Jackson, Tennessee Department of Education | John Zarian, The National Commission for the Certification of Crane Operators | Phil Dickison, The National Council of State Boards of Nursing | John Fremer, Caveon Test Security
Panel Presentation

As is often observed, attaining and maintaining test security in a high-stakes testing program requires many skills that must be consistently applied. Test Program Managers tend to be resourceful at drawing on their staff, Advisory Groups, and professionals in their discipline. Many programs face test security challenges that lead them to also reach out to other trained and experienced external professionals to build and maintain the test security aspects of their programs. In this session, participants from a variety of high-stakes programs will review the ways they drew on external expertise to supplement their program staff. Each presenter will draw on their experiences and the knowledge they have of best practices across testing programs to make recommendations regarding the use of external contributors. They will provide perspectives about strategies that have proved effective for them, as well as cautions about pitfalls to be avoided.

12:00 - 1:00

SALON F

LUNCH

GENERAL INFORMATION

CONFERENCE URL

conferenceontestsecurity.org

COTS SOCIAL MEDIA

#COTS2019

Twitter: https://twitter.com/COTS2019

FOOD ALLERGIES AND DIETARY RESTRICTIONS

All food allergies and dietary restrictions identified during the registration process have been communicated to the catering staff. If at any point you should have a question about the food selection, please speak with a member from Marriott Biscayne Bay, or stop by the registration/help desk.

LOCAL CONTACTS

Because it isn't possible to list every incredible event, restaurant, and attraction that beautiful sunshine Miami has to offer, please look for attendees with the yellow "local contact" ribbon on their name badge. They are familiar with the area and would be happy to help. You can also stop by the registration/help desk for additional resources.

THINGS TO DO IN MIAMI

https://www.miamiandbeaches.com/

https://www.tripadvisor.com/Home-g34438

https://themiamiguide.com/

SESSION INFORMATION

YOU CAN LOOK FORWARD TO THESE PRESENTATION FORMATS DURING COTS 2018

Standard Presentation: 60 or 90-minute session with 3–5 presentations on related topics. Each presentation in the session will be 15–20 minutes long.

Coordinated Symposium: Three to five separate research presentations, all focused on a common theme. One of the presentations may consist of a discussion, analysis, and/or contextualization of another session or sessions. All symposia will occupy either a 60 or 90-minute time slot.

Panel Presentation: Two to five individuals discussing different aspects of a common theme with each other and the audience. All panel presentations will occupy either a 60 or 90-minute time slot.

Demonstration: 60 or 90-minute session in which presenters demonstrate a technique or method related to a core aspect of test security.

Workshop: A 1½-hour or 3½-hour deep dive into specific topics or key security resources, which provides attendees with an opportunity to gain hands-on experience and collaborate with presenters and other attendees.

Facilitated Roundtable: 60 or 90-minute session that promotes networking and invites audience engagement around an important test security topic. Session will begin with a short, informal presentation to frame a conversation, followed by a free-flowing dialogue among audience members.

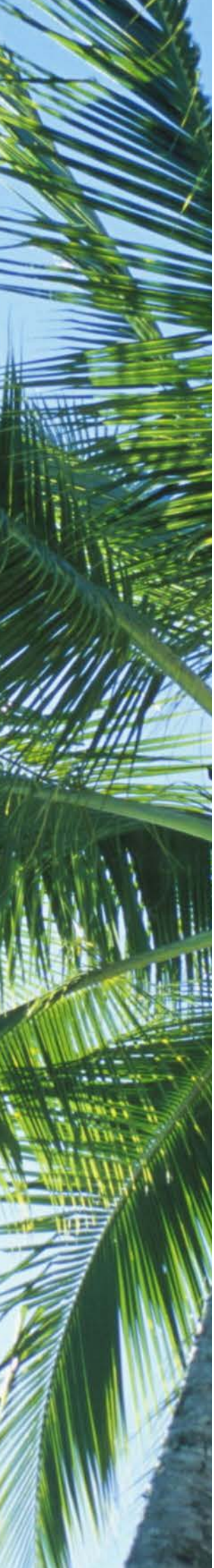
Poster Presentation: 60-minute session to include multiple posters on various topics. Each presenter will prepare a poster to fit on a 4' x 6' poster board. Poster presentations are informal and involve one-on-one interactions with many attendees.

NOTES

NOTES

NOTES

NOTES



SAVE THE DATE

THE CONFERENCE ON TEST SECURITY

OCTOBER 2020

PRINCETON, NJ

HOSTED BY
ETS

THANK YOU!

HOSTS

UNIVERSITY OF MIAMI

UNIVERSITY OF MIAMI
SCHOOL of EDUCATION & HUMAN DEVELOPMENT

CO-HOSTS

Pearson VUE

ACT®

ascend LEARNING

caveon™ Test Security

cisco

proctorU

ProctorioPro

FRIENDS

Alpine Testing Solutions

examity® BETTER TEST INTEGRITY.

NCSBN Leading Regulatory Excellence

psi

NCME national council on measurement in education

RENAISSANCE®

CollegeBoard

PROMETRIC®

National College Testing Association NCTA